

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097059 A2

(51) International Patent Classification⁷: C12N

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

(71) Applicant (*for all designated States except US*): CHROMOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081 Loughheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): PERKINS, Edward [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). FLEMING, Elena [CA/CA]; 248 E 18th, North Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). SHELLARD, Joan [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

Published:

— without international search report and to be republished upon receipt of that report

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.



WO 02/097059 A2

1-

CHROMOSOME-BASED PLATFORMS**RELATED APPLICATIONS**

Benefit of priority to U.S. provisional application Serial No. 60/294,758, filed May 30, 2001, to Perkins, *et al.*, entitled "CHROMOSOME-BASED PLATFORMS" and to U.S. provisional application Serial No. 60/366,891, filed March 21, 2002, to Perkins, *et al.*, entitled

5 "CHROMOSOME-BASED PLATFORMS" is claimed. Where permitted, the subject matter of which are herein incorporated by reference in their entirety.

This application is related to Provisional Application No. 60/294,687, filed May 30, 2001, by CARL PEREZ AND STEVEN

10 **FABIJANSKI** entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES* and to U.S. Provisional Application No. 60/296,329, filed June 4, 2001, by CARL PEREZ AND STEVEN FABIJANSKI entitled

15 *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES*. This application also is related to U.S. Provisional Application No. 60/294,758, filed May 30, 2001, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS* and to U.S. Provisional Application No. 60/366,891, filed March 21, 2002, by EDWARD PERKINS *et al.*, entitled

20 *CHROMOSOME-BASED PLATFORMS*. This application is also related to U.S. application Serial Nos. (attorney dkt nos. 24601-419 and 419PC), filed on the same day herewith, entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS OF PREPARING PLANT ARTIFICIAL CHROMOSOMES* to Perez *et al.*.

25 . This application is related to U.S. application Serial No. 08/695,191, filed August 7, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*

-2-

METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES, now U.S. Patent No. 6,025,155. This application is also related to U.S. application Serial No. 08/682,080, filed July 15, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF*

5 *AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,077,697. This application is also related U.S. application Serial No. 08/629,822, filed April 10, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL*

10 *CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned), and is also related to copending U.S. application Serial No. 09/096,648, filed June 12, 1998, by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL*

15 *CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* and to U.S. application Serial No. 09/835,682, April 10, 1997 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*

20 *METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned). This application is also related to copending U.S. application Serial No. 09/724,726, filed November 28, 2000, U.S. application Serial No. 09/724,872, filed November 28, 2000, U.S. application Serial No. 09/724,693, filed November 28, 2000, U.S. application Serial No. 09/799,462, filed March 5, 2001, U.S. application Serial No. 09/836,911, filed April 17, 2001, and U.S. application Serial No. 10/125,767, filed April 17, 2002, each of which is by GYULA

25 HADLACZKY and ALADAR SZALAY, and is entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also related to International PCT application No. WO 97/40183. Where permitted the

-3-

subject matter of each of these provisional applications, international applications, and applications is incorporated by reference in its entirety.

FIELD OF INVENTION

Artificial chromosomes, including *ACes*, that have been engineered
5 to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome.

BACKGROUND

Artificial chromosomes

10 A variety of artificial chromosomes for use in plants and animals, particularly higher plants and animals are available. In particular, U.S. Patent Nos. 6,025,155 and 6,077,697 provide heterochromatic artificial chromosomes designated therein as satellite artificial chromosomes
15 (*SATACs*) and now designated artificial chromosome expression systems (*ACes*). These chromosomes are prepared by introducing heterologous DNA into a selected plant or animal cell under conditions that result in integration into a region of the chromosome that leads to an amplification event resulting in production of a dicentric chromosome. Subsequent treatment and growth of cells with dicentric chromosomes, including
20 further amplifications, ultimately results in the artificial chromosomes provided therein. In order to introduce a desired heterologous gene (or a plurality of heterologous genes) into the artificial chromosome, the process is repeated introducing the desired heterologous genes and nucleic acids in the initial targeting step. This process is time consuming
25 and tedious. Hence, more tractable and efficient methods for introducing heterologous nucleic acid molecules into artificial chromosomes, particularly *ACes*, are needed.

-4-

Therefore, it is an object herein to provide engineered artificial chromosomes that permit tractable, efficient and rational engineering of artificial chromosomes.

SUMMARY OF THE INVENTION

5 Provided herein are artificial chromosomes that permit tractable, efficient and rational engineering thereof. In particular, the artificial chromosomes provided herein contain one or a plurality of loci (sites) for site-specific, recombination-directed integration of DNA. Thus, provided herein are platform artificial chromosome expression systems ("platform
10 *ACes*") containing single or multiple site-specific, recombination sites. The artificial chromosomes and *ACes* artificial chromosomes include plant and animal chromosomes. Any recombinase system that effects site-specific recombination is contemplated for use herein.

 In one embodiment, chromosomes, including platform *ACes*, are
15 provided that contain one or more lambda *att* sites designed for recombination-directed integration in the presence of lambda integrase, and that are mutated so that they do not require additional factors. Methods for preparing such chromosomes, vectors for use in the methods, and uses of the resulting chromosomes are also provided.

20 Platform *ACes* containing the recombination site(s) and methods for introducing heterologous nucleic acid into such sites and vectors therefor, are provided.

 Also provided herein is a bacteriophage lambda (λ) integrase site-specific recombination system.

25 Methods using recombinase mediated recombination target gene expression vectors and/or genes for insertion thereof into platform chromosomes and the resulting chromosomes are provided.

 Combinations and kits containing the combinations of vectors encoding a recombinase and integrase and primers for introduction of the

-5-

site recognized thereby are also provided. The kits optionally include instructions for performing site-directed integration or preparation of *ACes* containing such sites.

Also provided herein are mammalian and plant cells comprising the
5 artificial chromosomes and *ACes* described herein. The cells can be nuclear donor cells, stem cells, such as a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell or an embryonic stem cell.

Also provide is a lambda-intR mutain comprising a glutamic acid to arginine change at position 174 of wild-type lambda-integrase3. Also
10 provided are transgenic animals and methods for producing a transgenic animal, comprising introducing a *ACes* into an embryonic cell, such as a stem cell or embryo. The *ACes* can comprise heterologous nucleic acid that encodes a therapeutic product. The transgenic animal can be a fish, insect, reptile, amphibians, arachnid or mammal. In certain embodiments,
15 the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

The platform *ACes*, including plant and animal *ACes*, such as MACs, provided herein can be introduced into cells, such as, but not
20 limited to, animal cells, including mammalian cells, and into plant cells. Hence plant cells that contain platform MACs, animal cells that contain platform PACs and other combinations of cells and platform *ACes* are provided.

DESCRIPTION OF FIGURES

25 FIGURE 1 provides a diagram depicting creation of an exemplary *ACes* artificial chromosome prepared using methods detailed in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183. In this exemplified embodiment, the nucleic acid is targeted to an acrocentric chromosome in an animal or plant, and the

-6-

heterologous nucleic acid includes a sequence-specific recombination site and marker genes.

FIGURE 2 provides a map of pWEPuro9K, which is a targeting vector derived from the vector pWE15 (GenBank Accession # X65279; SEQ ID No. 31). Plasmid pWE15 was modified by replacing the *Sa*I (Klenow filled)/*Sma*I neomycin resistance encoding fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance-encoding fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc., Palo Alto, CA; GenBank Accession no. U07648; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (see Example 1, see, also Csonka *et al.* (2000) *Journal of Cell Science* 113:3207-32161; and SEQ ID NO: 118), containing a portion of the mouse rDNA region, was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K.

FIGURE 3 depicts construction of an ACes platform chromosome with a single recombination site, such as loxP sites or an *att*P or *att*B site. This platform ACes chromosome is an exemplary artificial chromosome with a single recombination site.

FIGURE 4 provides a map of plasmid pSV40-193attPsensePur.

FIGURE 5 depicts a method for formation of a chromosome platform with multiple recombination integration sites, such as *att*P sites.

FIGURE 6 sets forth the sequences of the core region of *att*P, *att*B, *att*L and *att*R (SEQ ID Nos. 33-36).

FIGURE 7 depicts insertional recombination of a vector encoding a marker gene, DsRed and an *att*B site with an artificial chromosome containing an *att*P site.

FIGURE 8 provides a map of plasmid pCXLamIntR (SEQ ID NO: 112), which includes the Lambda integrase (E174R)-encoding nucleic acid.

-7-

FIGURE 9 diagrammatically summarizes the platform technology; marker 1 permits selection of the artificial chromosomes containing the integration site; marker 2, which is promoterless in the target gene expression vector, permits selection of recombinants. Upon
5 recombination with the platform marker 2 is expressed under the control of a promoter resident on the platform.

FIGURE 10 provides the vector map for the plasmid p18attBZEO-5'6XHS4eGFP (SEQ ID NO: 116).

FIGURE 11 provides the vector map for the plasmid p18attBZEO-
10 3'6XHS4eGFP (SEQ ID NO: 115).

FIGURE 12 provides the vector map for the plasmid p18attBZEO-(6XHS4)2eGFP (SEQ ID NO: 110).

FIGURES 13 AND 14 depict the integration of a PCR product by site-specific recombination as set forth in Example 8.
15 FIGURE 15 provides the vector map for the plasmid pPACrDNA as set forth in Example 9.A.

DETAILED DESCRIPTION OF THE INVENTION

A. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used
20 herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference
25 in their entirety. Where reference is made to a URL or other such identifier or address, it is understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

-8-

As used herein, nucleic acid refers to single-stranded and/or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique and of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleotides long.

As used herein, DNA is meant to include all types and sizes of DNA molecules including cDNA, plasmids and DNA including modified nucleotides and nucleotide analogs.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified-nucleotides, such as, but are not limited to, phosphorothioate nucleotides and deazapurine nucleotides and other nucleotide analogs.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations and/or in amounts in a genome or cell that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not

-9-

normally produced by the cell in which it is expressed. Any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Heterologous DNA and RNA may also encode RNA or proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes.

Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of modification of the endogenous genes or for production of an encoded protein. For example, a heterologous or foreign gene may be isolated from a different species than that of the host genome, or alternatively, may be isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Other examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers traits including, but not limited to, herbicide, insect, or disease resistance; traits, including, but not limited to, oil quality or carbohydrate composition. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

As used herein, operative linkage or operative association, or grammatical variations thereof, of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the

-10-

promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA.

In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak (1991) *J. Biol. Chem.* 266:19867-19870) can be inserted immediately 5' of the start codon and may enhance expression.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, preferably under moderate or high stringency conditions, forming a stable duplex. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. The longer the hybridizing nucleic acid, the more base mismatches it can contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

As used herein, regulatory molecule refers to a polymer of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or a polypeptide that is capable of enhancing or inhibiting expression of a gene.

As used herein, recognition sequences are particular sequences of nucleotides that a protein, DNA, or RNA molecule, or combinations thereof, (such as, but not limited to, a restriction endonuclease, a modification methylase and a recombinase) recognizes and binds. For example, a recognition sequence for Cre recombinase (see, *e.g.*, SEQ ID NO:58) is a 34 base pair sequence containing two 13 base pair inverted

-11-

repeats (serving as the recombinase binding sites) flanking an 8 base pair core and designated loxP (see, *e.g.*, Sauer (1994) *Current Opinion in Biotechnology* 5:521-527). Other examples of recognition sequences, include, but are not limited to, attB and attP, attR and attL and others (see, *e.g.*, SEQ ID Nos. 8, 41-56 and 72), that are recognized by the recombina-
5 se enzyme Integrase (see, SEQ ID Nos. 37 and 38 for the nucleotide and encoded amino acid sequences of an exemplary lambda phage integrase).

The recombination site designated attB is an approximately 33 base pair sequence containing two 9 base pair core-type Int binding sites and a
10 7 base pair overlap region; attP (SEQ ID No. 72) is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis (see, *e.g.*, Landy (1993) *Current Opinion in Biotechnology* 3:699-707) see,
15 *e.g.*, SEQ ID Nos. 8 and 72).

As used herein, a recombinase is an enzyme that catalyzes the exchange of DNA segments at specific recombination sites. An integrase herein refers to a recombinase that is a member of the lambda (λ) integrase family.

20 As used herein, recombination proteins include excisive proteins, integrative proteins, enzymes, co-factors and associated proteins that are involved in recombination reactions using one or more recombination sites (see, Landy (1993) *Current Opinion in Biotechnology* 3:699-707). The recombination proteins used herein can be delivered to a cell via an
25 expression cassette on an appropriate vector, such as a plasmid, and the like. In other embodiments, the recombination proteins can be delivered to a cell in protein form in the same reaction mixture used to deliver the desired nucleic acid, such as a platform ACes, donor target vectors, and the like.

-12-

As used herein the expression "lox site" means a sequence of nucleotides at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination event. A LoxP site is a 34 base pair nucleotide sequence from bacteriophage P1 (see, 5 *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402). The LoxP site contains two 13 base pair inverted repeats separated by an 8 base pair spacer region as follows: (SEQ ID NO. 57):

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT

E. coli DH5Δlac and yeast strain BSY23 transformed with plasmid pBS44 10 carrying two loxP sites connected with a LEU2 gene are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53254 and ATCC 20773, respectively. The lox sites can be isolated from plasmid pBS44 with restriction enzymes *EcoRI* and *SalI*, or *XhoI* and *BamHI*. In addition, a preselected DNA segment can be inserted 15 into pBS44 at either the *SalI* or *BamHI* restriction enzyme sites. Other lox sites include, but are not limited to, LoxB, LoxL, LoxC2 and LoxR sites, which are nucleotide sequences isolated from *E. coli* (see, *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398). Lox sites can also be produced by a variety of synthetic techniques (see, *e.g.*, Ito *et al.* (1982) 20 *Nuc. Acid Res.* 10:1755 and Ogilvie *et al.* (1981) *Science* 270:270).

As used herein, the expression "cre gene" means a sequence of nucleotides that encodes a gene product that effects site-specific recombination of DNA in eukaryotic cells at lox sites. One cre gene can be isolated from bacteriophage P1 (see, *e.g.*, Abremski *et al.* (1983) *Cell* 25 32:1301-1311). *E. coli* DH1 and yeast strain BSY90 transformed with plasmid pBS39 carrying a cre gene isolated from bacteriophage P1 and a GAL1 regulatory nucleotide sequence are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53255

-13-

and ATCC 20772, respectively. The cre gene can be isolated from plasmid pBS39 with restriction enzymes *XhoI* and *SaII*.

As used herein, site-specific recombination refers to site-specific recombination that is effected between two specific sites on a single
5 nucleic acid molecule or between two different molecules that requires the presence of an exogenous protein, such as an integrase or recombinase.

For example, Cre-lox site-specific recombination can include the following three events:

- 10 a. deletion of a pre-selected DNA segment flanked by lox sites;
- b. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites; and
- c. reciprocal exchange of DNA segments proximate to
15 lox sites located on different DNA molecules.

This reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. DNA segment refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source.

- 20 Since the lox site is an asymmetrical nucleotide sequence, two lox sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA segment located between the two lox sites and a connection between the resulting ends of the original
25 DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox

-14-

sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the gene product of the cre gene. Thus, the Cre-lox system can be used to specifically delete, invert, or
5 insert DNA. The precise event is controlled by the orientation of lox DNA sequences, in *cis* the lox sequences direct the *Cre* recombinase to either delete (lox sequences in direct orientation) or invert (lox sequences in inverted orientation) DNA flanked by the sequences, while *in trans* the lox sequences can direct a homologous recombination event resulting in the
10 insertion of a recombinant DNA.

As used herein, a chromosome is a nucleic acid molecule, and associated proteins, that is capable of replication and segregation within a cell upon cell division. Typically, a chromosome contains a centromeric region, replication origins, telomeric regions and a region of nucleic acid
15 between the centromeric and telomeric regions.

As used herein, a centromere is any nucleic acid sequence that confers an ability to segregate to daughter cells through cell division. A centromere may confer stable segregation of a nucleic acid sequence, including an artificial chromosome containing the centromere, through
20 mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A particular centromere is not necessarily derived from the same species in which it is introduced, but has the ability to promote DNA segregation in cells of that species.

As used herein, euchromatin and heterochromatin have their
25 recognized meanings. Euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to be transcriptionally inactive. Highly repetitive DNA sequences (satellite DNA) are usually located in regions of the heterochromatin surrounding

-15-

the centromere (pericentric or pericentromeric heterochromatin).

Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

5 As used herein, an acrocentric chromosome refers to a chromosome with arms of unequal length.

 As used herein, endogenous chromosomes refer to genomic chromosomes as found in a cell prior to generation or introduction of an artificial chromosome.

10 As used herein, artificial chromosomes are nucleic acid molecules, typically DNA, that stably replicate and segregate alongside endogenous chromosomes in cells and have the capacity to accommodate and express heterologous genes contained therein. It has the capacity to act as a gene delivery vehicle by accommodating and expressing foreign genes
15 contained therein. A mammalian artificial chromosome (MAC) refers to chromosomes that have an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes that include centromeres that function in plant, insect and avian cells, respectively. A human artificial
20 chromosome (HAC) refers to chromosomes that include centromeres that function in human cells. For exemplary artificial chromosomes, see, *e.g.*, U.S. Patent Nos. 6,025,155; 6,077,697; 5,288,625; 5,712,134; 5,695,967; 5,869,294; 5,891,691 and 5,721,118 and published International PCT application Nos, WO 97/40183 and WO 98/08964.
25 Artificial chromosomes include those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), minichromosomes that contain a *de novo* centromere (see, U.S. Patent

-16-

Nos. 5,712,134, 5,891,691 and 5,288,625), artificial chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA and *in vitro* assembled artificial chromosomes (see, copending U.S. 5 provisional application Serial No. 60/294,687, filed on May 30, 2001).

As used herein, the term "satellite DNA-based artificial chromosome (SATAC)" is interchangeable with the term "artificial chromosome expression system (ACes)". These artificial chromosomes (ACes) include those that are substantially all neutral non-coding 10 sequences (heterochromatin) except for foreign heterologous, typically gene-encoding nucleic acid, that is interspersed within the heterochromatin for the expression therein (see U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183), or that is in a single locus as provided herein. Also included 15 are ACes that may include euchromatin and that result from the process described in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183 and outlined herein. The delineating structural feature is the presence of repeating units, that are generally predominantly heterochromatin. The precise structure of the ACes will 20 depend upon the structure of the chromosome in which the initial amplification event occurs; all share the common feature of including a defined pattern of repeating units. Generally ACes have more heterochromatin than euchromatin. Foreign nucleic acid molecules (heterologous genes) contained in these artificial chromosome expression 25 systems can include any nucleic acid whose expression is of interest in a particular host cell. Such foreign nucleic acid molecules, include, but are not limited to, nucleic acid that encodes traceable marker proteins (reporter genes), such as fluorescent proteins, such as green, blue or red fluorescent proteins (GFP, BFP and RFP, respectively), other reporter

-17-

genes, such as β -galactosidase and proteins that confer drug resistance, such as a gene encoding hygromycin-resistance. Other examples of heterologous nucleic acid molecules include, but are not limited to, DNA that encodes therapeutically effective substances, such as anti-cancer
5 agents, enzymes and hormones, DNA that encodes other types of proteins, such as antibodies, and DNA that encodes RNA molecules (such as antisense or siRNA molecules) that are not translated into proteins.

As used herein, an artificial chromosome platform, also referred to
10 herein as a "platform ACes" or "ACes platform", refers to an artificial chromosome that has been engineered to include one or more sites for site-specific, recombination-directed integration. In particular, ACes that are so-engineered are provided. Any sites, including but not limited to any described herein, that are suitable for such integration are
15 contemplated. Plant and animal platform ACes are provided. Among the ACes contemplated herein are those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), artificial
20 chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA resulting from an amplification event depicted in the referenced patent and herein. Included among the ACes for use in generating platforms, are artificial chromosomes that introduce and
25 express heterologous nucleic acids in plants (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001). These include artificial chromosomes that have a centromere derived from a plant, and, also, artificial chromosomes that have centromeres that may be derived from other organisms but that function in plants.

-18-

As used herein a "reporter ACes" refers to a an ACes that comprises one or a plurality of reporter constructs, where the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

5 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of substantially similar or identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

10 As used herein, amplification-based artificial chromosomes are artificial chromosomes derived from natural or endogenous chromosomes by virtue of an amplification event, such as one initiated by introduction of heterologous nucleic acid into rDNA in a chromosome. As a result of such an event, chromosomes and fragments thereof exhibiting segmented
15 or repeating patterns arise. Artificial chromosomes can be formed from these chromosomes and fragments. Hence, amplification-based artificial chromosomes refer to engineered chromosomes that exhibit an ordered segmentation that is not observed in naturally occurring chromosomes and that distinguishes them from naturally occurring chromosomes. The
20 segmentation, which can be visualized using a variety of chromosome analysis techniques known to those of skill in the art, correlates with the structure of these artificial chromosomes. In addition to containing one or more centromeres, the amplification-based artificial chromosomes, throughout the region or regions of segmentation are predominantly made
25 up of nucleic acid units also referred to as "amplicons", that is (are) repeated in the region and that have a similar gross structure. Repeats of an amplicon tend to be of similar size and share some common nucleic acid sequences. For example, each repeat of an amplicon may contain a replication site involved in amplification of chromosome segments and/or

-19-

some heterologous nucleic acid that was utilized in the initial production of the artificial chromosome. Typically, the repeating units are substantially similar in nucleic acid composition and may be nearly identical.

5 The amplification-based artificial chromosomes differ depending on the chromosomal region that has undergone amplification in the process of artificial chromosome formation. The structures of the resulting chromosomes can vary depending upon the initiating event and/or the conditions under which the heterologous nucleic acid is introduced,
10 including modification to the endogenous chromosomes. For example, in some of the artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of heterochromatic DNA. In other artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of euchromatic DNA or may
15 be made up of similar amounts of heterochromatic and euchromatic DNA.

As used herein an amplicon is a repeated nucleic acid unit. In some of the artificial chromosomes described herein, an amplicon may contain a set of inverted repeats of a megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to
20 some of the predominantly heterochromatic artificial chromosomes, the megareplicon can contain a set of tandem DNA blocks (*e.g.*, ~7.5 Mb DNA blocks) each containing satellite DNA flanked by non-satellite DNA or may be made up of substantially rDNA. Contained within the megareplicon is a primary replication site, referred to as the
25 megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller (*e.g.*, 50-300 kb) secondary replicons.

-20-

In artificial chromosomes, such as those provided U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183, the megareplicon is defined by two tandem blocks (~7.5 Mb DNA blocks in the chromosomes provided therein). Within each artificial chromosome or among a population thereof, each amplicon has the same gross structure but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the artificial chromosomes or their overall structure as described herein.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating artificial chromosomes provided herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other cellular events involving recombination (*e.g.*, DNA repair). Such regions include regions of the chromosome that contain tandem repeats, such as satellite DNA, rDNA, and other such sequences.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments is a replicable chromosome. If one of the chromosomes undergoes amplification of primarily euchromatic DNA to produce a fully functional chromosome that is predominantly (at least more than 50%) euchromatin, it is a minichromosome. The remaining chromosome is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification,

-21-

whereby heterochromatin (such as, for example, satellite DNA) is amplified and a euchromatic portion (such as, for example, an arm) remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a predominantly heterochromatic artificial chromosome. Predominantly heterochromatic artificial chromosomes can be produced from other partially heterochromatic artificial chromosomes by culturing the cell containing such chromosomes under conditions such as BrdU treatment that destabilize the chromosome and/or growth under selective conditions so that a predominantly heterochromatic artificial chromosome is produced. For purposes herein, it is understood that the artificial chromosomes may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous DNA. Typically, artificial chromosomes appear after about 5 to about 60, or about 5 to about 55, or about 10 to about 55 or about 25 to about 55 or about 35 to about 55 cell doublings after initiation of artificial chromosome generation, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, an artificial chromosome that is predominantly heterochromatic (*i.e.*, containing more heterochromatin than euchromatin, typically more than about 50%, more than about 70%, or more than about 90% heterochromatin) may be produced by introducing nucleic acid molecules into cells, such as, for example, animal or plant cells, and selecting cells that contain a predominantly heterochromatic artificial chromosome. Any nucleic acid may be introduced into cells in such methods of producing the artificial chromosomes. For example, the nucleic acid may contain a selectable marker and/or optionally a sequence that targets nucleic acid to the pericentric, heterochromatic region of a chromosome, such as in the short arm of acrocentric chromosomes and

-22-

nucleolar organizing regions. Targeting sequences include, but are not limited to, lambda phage DNA and rDNA for production of predominantly heterochromatic artificial chromosomes in eukaryotic cells.

After introducing the nucleic acid into cells, a cell containing a
5 predominantly heterochromatic artificial chromosome is selected. Such
cells may be identified using a variety of procedures. For example,
repeating units of heterochromatic DNA of these chromosomes may be
discerned by G-banding and/or fluorescence in situ hybridization (FISH)
techniques. Prior to such analyses, the cells to be analyzed may be
10 enriched with artificial chromosome-containing cells by sorting the cells
on the basis of the presence of a selectable marker, such as a reporter
protein, or by growing (culturing) the cells under selective conditions. It
is also possible, after introduction of nucleic acids into cells, to select
cells that have a multicentric, typically dicentric, chromosome, a formerly
15 multicentric (typically dicentric) chromosome and/or various
heterochromatic structures, such as a megachromosome and a sausage
chromosome, that contain a centromere and are predominantly
heterochromatic and to treat them such that desired artificial
chromosomes are produced. Cells containing a new chromosome are
20 selected. Conditions for generation of a desired structure include, but are
not limited to, further growth under selective conditions, introduction of
additional nucleic acid molecules and/or growth under selective conditions
and treatment with destabilizing agents, and other such methods (see
International PCT application No. WO 97/40183 and U.S. Patent Nos.
25 6,025,155 and 6,077,697).

As used herein, a "selectable marker" is a nucleic acid segment,
generally DNA, that allows one to select for or against a molecule or a
cell that contains it, often under particular conditions. These markers can
encode an activity, such as, but not limited to, production of RNA,

-23-

peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds and compositions. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products that provide resistance against otherwise
5 toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be identified, such as phenotypic markers,
10 including β -galactosidase, red, blue and/or green fluorescent proteins (FPs), and cell surface proteins; (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense
15 oligonucleotides or siRNA molecules for use in RNA interference); (7) nucleic acid segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) nucleic acid segments that can be used to isolate a desired molecule (e.g. specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be
20 otherwise non-functional, such as for PCR amplification of subpopulations of molecules; and/or (10) nucleic acid segments, which when absent, directly or indirectly confer sensitivity to particular compounds. Thus, for example, selectable markers include nucleic acids encoding fluorescent proteins, such as green fluorescent proteins, β -galactosidase and other
25 readily detectable proteins, such as chromogenic proteins or proteins capable of being bound by an antibody and FACs sorted. Selectable markers such as these, which are not required for cell survival and/or proliferation in the presence of a selection agent, are also referred to herein as reporter molecules. Other selectable markers, *e.g.*, the

-24-

neomycin phosphotransferase gene, provide for isolation and identification of cells containing them by conferring properties on the cells that make them resistant to an agent, *e.g.*, a drug such as an antibiotic, that inhibits proliferation of cells that do not contain the marker.

- 5 As another example, interference of gene expression by double stranded RNA has been shown in *Caenorhabditis elegans*, plants, *Drosophila*, protozoans and mammals. This method is known as RNA interference (RNAi) and utilizes short, double-stranded RNA molecules (siRNAs). The siRNAs are generally composed of a 19-22bp double-
- 10 stranded RNA stem, a loop region and a 1-4 bp overhang on the 3' end. The reduction of gene expression has been accomplished by direct introduction of the siRNAs into the cell (Harborth J et al., 2001, J Cell Sci 114(pt 24):4557-65) as well as the introduction of DNA encoding and expressing the siRNA molecule. The encoded siRNA molecules are under
- 15 the regulation of an RNA polymerase III promoter (see, *e.g.*, Yu et al., 2002, Proc Natl Acad Sci USA 99(9):6047-52; Brummelkamp et al., 2002, Science 296(5567):550-3; Miyagishi et al., 2002, Nat Biotechnol 20(5):497-500; and the like). In certain embodiments, RNAi in mammalian cells may have advantages over other therapeutic methods.
- 20 For example, producing siRNA molecules that block viral genetic activities in infected cells may reduce the effects of the virus. Platform ACes provided herein encoding siRNA molecule(s) are an additional utilization of the platform ACes technology. The platform ACes could be engineered to encode one or more siRNA molecules to create gene "knockdowns". In
- 25 one embodiment, a platform ACes can engineered to encode both the siRNA molecule and a replacement gene. For example, a mouse model or cell culture system could be generated using a platform ACes that has a knockdown of the endogenous mouse gene, by siRNA, and the human gene homolog expressing in place of the mouse gene. The placement of

-25-

siRNA encoding sequences under the regulation of a regulatable or inducible promoter would allow one to temporally and/or spatially control the knockdown effect of the corresponding gene.

As used herein, a reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein. Generally reporter genes are readily detectable. Examples of reporter genes include, but are not limited to nucleic acid encoding a fluorescent protein, CAT (chloramphenicol acetyl transferase) (Alton *et al.* (1979) *Nature* 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet *et al.* (1987) *Mol. Cell. Biol.* 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:4154-4158; Baldwin *et al.* (1984) *Biochemistry* 23:3663-3667); and alkaline phosphatase (Toh *et al.* (1989) *Eur. J. Biochem.* 182:231-238, Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101).

As used herein, growth under selective conditions means growth of a cell under conditions that require expression of a selectable marker for survival.

As used herein, an agent that destabilizes a chromosome is any agent known by those skilled in the art to enhance amplification events, and/or mutations. Such agents, which include BrdU, are well known to those skilled in the art.

In order to generate an artificial chromosome containing a particular heterologous nucleic acid of interest, it is possible to include the nucleic acid in the nucleic acid that is being introduced into cells to initiate production of the artificial chromosome. Thus, for example, a nucleic acid can be introduced into a cell along with nucleic acid encoding a selectable marker and/or a nucleic acid that targets to a heterochromatic region of a chromosome. For introducing a heterologous nucleic acid into

-26-

the cell, it can be included in a fragment that includes a selectable marker or as part of a separate nucleic acid fragment and introduced into the cell with a selectable marker during the process of generating the artificial chromosomes. Alternatively, heterologous nucleic acid can be introduced
5 into an artificial chromosome at a later time after the initial generation of the artificial chromosome.

As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome that contains more euchromatic than heterochromatic DNA. For purposes herein, the
10 minichromosome contains a *de novo* centromere (e.g., a neocentromere). In some embodiments, for example, the minichromosome contains a centromere that replicates in animals, e.g., a mammalian centromere or in plants, e.g., a plant centromere.

As used herein, *in vitro* assembled artificial chromosomes or
15 synthetic chromosomes can be either more euchromatic than heterochromatic or more heterochromatic than euchromatic and are produced by joining essential components of a chromosome *in vitro*. These components include at least a centromere, a megareplicator, a telomere and optionally secondary origins of replication.

As used herein, *in vitro* assembled plant or animal artificial
20 chromosomes are produced by joining essential components (at least the centromere, telomere(s), megareplicator and optional secondary origins of replication) that function in plants or animals. In particular embodiments, the megareplicator contains sequences of rDNA, particularly plant or
25 animal rDNA.

As used herein, a plant is a eukaryotic organism that contains, in addition to a nucleus and mitochondria, chloroplasts capable of carrying out photosynthesis. A plant can be unicellular or multicellular and can contain multiple tissues and/or organs. Plants can reproduce sexually or

-27-

asexually and can be perennial or annual in growth. Plants can also be terrestrial or aquatic. The term "plant" includes a whole plant, plant cell, plant protoplast, plant calli, plant seed, plant organ, plant tissue, and other parts of a whole plant.

- 5 As used herein, stable maintenance of chromosomes occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their
10 structure during cell culturing, suffering no unintended intrachromosomal or interchromosomal rearrangements.

- As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere in a chromosome as a result of incorporation of a heterologous nucleic acid fragment using the methods
15 herein.

- As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell
division.

- 20 As used herein, ribosomal RNA (rRNA) is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes (i.e., genes which encode rRNA)
25 per haploid genome are spread out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA (rDNA, which is DNA containing sequences that encode rRNA) has been verified on at least 11 pairs out of 20 mouse chromosomes (chromosomes 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, and 19)

-28-

(see e.g., Rowe *et al.* (1996) *Mamm. Genome* 7:886-889 and Johnson *et al.* (1993) *Mamm. Genome* 4:49-52). In *Arabidopsis thaliana* the presence of rDNA has been verified on chromosomes 2 and 4 (18S, 5.8S, and 25S rDNA) and on chromosomes 3,4, and 5 (5S rDNA)(see The
5 *Arabidopsis* Genome Initiative (2000) *Nature* 408:796-815). In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA
10 which can vary in length and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of
15 ribosome production within the cell nucleus.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of heterochromatin. Megachromosomes are made up of an array of repeated amplicons that contain two inverted megareplicons bordered by
20 introduced heterologous DNA (see, *e.g.*, Figure 3 of U.S. Patent No. 6,077,697 for a schematic drawing of a megachromosome). For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes (about 90 to 120 or 150 Mb), dwarf
25 megachromosomes (~ 150-200 Mb), and a micro-megachromosome (~ 50-90 Mb, typically 50-60 Mb). For purposes herein, the term

-29-

megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments (amplicons) that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

5 As used herein, gene therapy involves the transfer or insertion of nucleic acid molecules into certain cells, which are also referred to as target cells, to produce specific products that are involved in preventing, curing, correcting, controlling or modulating diseases, disorders and deleterious conditions. The nucleic acid is introduced into the selected
10 target cells in a manner such that the nucleic acid is expressed and a product encoded thereby is produced. Alternatively, the nucleic acid may in some manner mediate expression of DNA that encodes a therapeutic product. This product may be a therapeutic compound, which is produced in therapeutically effective amounts or at a therapeutically
15 useful time. It may also encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Expression of the nucleic acid by the target cells within an organism afflicted with a disease or disorder thereby provides for modulation of the disease or disorder. The nucleic acid encoding the
20 therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

For use in gene therapy, cells can be transfected *in vitro*, followed by introduction of the transfected cells into an organism. This is often
25 referred to as *ex vivo* gene therapy. Alternatively, the cells can be transfected directly *in vivo* within an organism.

As used herein, therapeutic agents include, but are not limited to, growth factors, antibodies, cytokines, such as tumor necrosis factors and interleukins, and cytotoxic agents and other agents disclosed herein and

-30-

known to those of skill in the art. Such agents include, but are not limited to, tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), erythropoietin (EPO), pro-coagulants such as tissue factor and tissue factor variants, pro-apoptotic agents such as FAS-ligand, fibroblast growth factors (FGF), nerve growth factor and other growth factors.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures the disease.

As used herein, transgenic plants and animals refer to plants and animals in which heterologous or foreign nucleic acid is expressed or in which the expression of a gene naturally present in the plant or animal has been altered by virtue of introduction of heterologous or foreign nucleic acid.

As used herein, IRES (internal ribosome entry site; see, *e.g.*, SEQ ID No. 27 and nucleotides 2736-3308 SEQ ID No. 28) refers to a region of a nucleic acid molecule, such as an mRNA molecule, that allows internal ribosome entry sufficient to initiate translation, which initiation can be detected in an assay for cap-independent translation (see, *e.g.*, U.S. Patent No. 6,171,821). The presence of an IRES within an mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated.

-31-

Internal ribosome entry site (IRES) elements were first identified in picornaviruses, which elements are considered the paradigm for cap-independent translation. The 5' UTRs of all picornaviruses are long and mediate translational initiation by directly recruiting and binding
5 ribosomes, thereby circumventing the initial cap-binding step. IRES elements are frequently found in viral mRNA, they are rare in non-viral mRNA. Among non-viral mRNA molecules that contain functional IRES elements in their respective 5' UTRs are those encoding immunoglobulin heavy chain binding protein (BiP) (Macejak *et al.* (1991) *Nature*
10 353:90-94); *Drosophila* Antennapedia (Oh *et al.* (1992) *Genes Dev.* 6:1643-1653); D. Ultrabithorax (Ye *et al.* (1997) *Mol. Cell Biol.* 17:1714-21); fibroblast growth factor 2 (Vagner *et al.* (1995) *Mol. Cell Biol.* 15:35-44); initiation factor eIF4G (Gan *et al.* (1998) *J. Biol. Chem.* 273:5006-5012); proto-oncogene c-myc (Nanbru *et al.* (1995) *J. Biol.*
15 *Chem.* 272:32061-32066; Stoneley (1998) *Oncogene* 16:423-428); IRES_H; from the 5'UTR of NRF1 gene (Oumard *et al.* (2000) *Mol. and Cell Biol.*, 20(8):2755-2759); and vascular endothelial growth factor (VEGF) (Stein *et al.* (1998) *Mol. Cell Biol.* 18:3112-9).

As used herein, a promoter, with respect to a region of DNA, refers
20 to a sequence of DNA that contains a sequence of bases that signals RNA polymerase to associate with the DNA and initiate transcription of RNA (such as pol II for mRNA) from a template strand of the DNA. A promoter thus generally regulates transcription of DNA into mRNA. A particular promoter provided herein is the Ferritin heavy chain promoter (excluding
25 the Iron Response Element, located in the 5'UTR), which was joined to the 37bp Fer-1 enhancer element. This promoter is set forth as SEQ ID NO:128. The endogenous Fer-1 enhancer element is located upstream of the Fer-1 promoter (e.g., a Fer-1 oligo was cloned proximal to the core promoter).

-32-

As used herein, isolated, substantially pure nucleic acid, such as, for example, DNA, refers to nucleic acid fragments purified according to standard techniques employed by those skilled in the art, such as that found in Sambrook *et al.* ((2001) Molecular Cloning: A Laboratory
5 Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition).

As used herein, expression refers to the transcription and/or translation of nucleic acid. For example, expression can be the transcription of a gene that may be transcribed into an RNA molecule,
10 such as a messenger RNA (mRNA) molecule. Expression may further include translation of an RNA molecule and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA. With respect to an antisense
15 construct, expression may refer to the transcription of the antisense DNA.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous nucleic acids into cells for either expression of the heterologous nucleic acid or for replication of the heterologous nucleic acid. Selection and use of such vectors and
20 plasmids are well within the level of skill of the art.

As used herein, transformation/transfection refers to the process by which nucleic acid is introduced into cells. The terms transfection and transformation refer to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences
25 are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by *Agrobacterium*-mediated transformation, protoplast transformation (including polyethylene glycol (PEG)-mediated transformation, electroporation, protoplast fusion, and microcell fusion), lipid-mediated delivery, liposomes, electroporation,

-33-

sonoporation, microinjection, particle bombardment and silicon carbide whisker-mediated transformation and combinations thereof (see, *e.g.*, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:169-177; Reich *et al.* (1986) *Biotechnology* 4:1001-1004; Klein *et al.* (1987) *Nature* 327:70-73; U.S. Patent No. 6,143,949; Paszkowski *et al.* (1989) in *Cell Culture and Somatic Cell Genetics of Plants, Vol. 6*, Molecular Biology of Plant Nuclear Genes, eds. Schell, J and Vasil, L.K. Academic Publishers, San Diego, California, p. 52-68; and Frame *et al.* (1994) *Plant J.* 6:941-948), direct uptake using calcium phosphate (CaPO₄; see, *e.g.*, Wigler *et al.* (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76:1373-1376), polyethylene glycol (PEG)-mediated DNA uptake, lipofection (see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327), microcell fusion (see, EXAMPLES, see, also Lambert (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5907-5911; U.S. Patent No. 5,396,767, Sawford *et al.* (1987) *Somatic Cell Mol. Genet.* 13:279-284; Dhar *et al.* (1984) *Somatic Cell Mol. Genet.* 10:547-559; and McNeill-Killary *et al.* (1995) *Meth. Enzymol.* 254:133-152), lipid-mediated carrier systems (see, *e.g.*, Teifel *et al.* (1995) *Biotechniques* 19:79-80; Albrecht *et al.* (1996) *Ann. Hematol.* 72:73-79; Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolch *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618) or other suitable method. Methods for delivery of ACes are described in copending U.S. application Serial No. 09/815,979. Successful transfection is generally recognized by detection of the presence of the heterologous nucleic acid within the transfected cell, such as, for example, any visualization of the heterologous nucleic acid or any indication of the operation of a vector within the host cell.

-34-

As used herein, "delivery," which is used interchangeably with "transfection," refers to the process by which exogenous nucleic acid molecules are transferred into a cell such that they are located inside the cell. Delivery of nucleic acids is a distinct process from expression of
5 nucleic acids.

As used herein, injected refers to the microinjection, such as by use of a small syringe, needle, or pipette, for injection of nucleic acid into a cell.

As used herein, substantially homologous DNA refers to DNA that
10 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids, with each other or a reference sequence, under specified conditions.

It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions,
15 hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 10, 14 or 16 or more nucleotides in a sequence that is complementary (or nearly complementary) to a substantially
20 contiguous sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with complementary base-pairing segments hybridize detectably to each other, departures from exact
25 complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids becomes larger, and as

-35-

conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence", if (a) both form a base-paired duplex with the same
5 segment, and (b) the melting temperatures of the two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of
10 nucleic acid duplexes are well known (see, *e.g.*, Meinkoth *et al.* (1984) *Anal. Biochem.* 138:267-284 and references cited therein).

As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes complementary or substantially complementary
15 sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt
20 concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

To be used as a hybridization probe, the nucleic acid is generally
25 rendered detectable by labeling it with a detectable moiety or label, such as ³²P, ³H and ¹⁴C, or by other means, including chemical labeling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

-36-

(via the biotin moieties) by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. (New York, NY). Any other label known to those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available (for culturing cells, extracting DNA, and hybridization assays), the quantity of DNA or RNA available as a source of the probe, the particular label and the means used to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques (see, e.g., Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory Press).

As used herein, conditions under which DNA molecules form stable hybrids are considered substantially homologous, and a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a polypeptide. By the term "substantially homologous" is meant having at least 75%, preferably 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970)), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e.,

-37-

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and
5 the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3)
10 no penalty for end gaps.

By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate
15 stringency or at high stringency all along the length of the nucleic acid of interest. Preferably the two molecules will hybridize under conditions of high stringency. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

20 Whether any two nucleic acid molecules have nucleotide sequences that are at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988).
25 Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine relative sequence identity.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

-38-

calculated using published techniques. (See, e.g.: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 5 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or

10 polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to

15 Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited

20 to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(11):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., et al., *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is

25 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of

-39-

100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

10 As used herein: stringency of hybridization in determining percentage mismatch encompass the following conditions or equivalent conditions thereto:

- 1) high stringency: 0.1 x SSPE or SSC, 0.1% SDS, 65°C
 - 2) medium stringency: 0.2 x SSPE or SSC, 0.1% SDS, 50°C
 - 15 3) low stringency: 1.0 x SSPE or SSC, 0.1% SDS, 50°C
- or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching. Equivalent conditions refer to conditions that select for substantially the same percentage of mismatch in the resulting hybrids. Additions of ingredients, such as formamide, Ficoll, and Denhardt's solution affect parameters such as the temperature under which the hybridization should be conducted and the rate of the reaction. Thus, hybridization in 5 X SSC, in 20% formamide at 42° C is substantially the same as the conditions recited above hybridization under conditions of low stringency. The recipes for 25 SSPE, SSC and Denhardt's and the preparation of deionized formamide are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Chapter 8; see, Sambrook *et al.*, vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions. It is understood that

-40-

equivalent stringencies may be achieved using alternative buffers, salts and temperatures. As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

As used herein, conservative amino acid substitutions, such as those set forth in Table 1, are those that do not eliminate biological activity. Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
20	Arg (R)	Lys, orn
	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
25	Gly (G)	Ala; Pro
	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nva
	Lys (K)	Arg; Gln; Glu
30	Met (M)	Leu; Tyr; Ile; NLe Val
	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
35	Trp (W)	Tyr
	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nva

-41-

Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results in more than one type of mRNA.

As used herein, a probe or primer based on a nucleotide sequence includes at least 10, 14, 16, 30 or 100 contiguous nucleotides from the reference nucleic acid molecule.

As used herein, recombinant production by using recombinant DNA methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a substrate, light is produced.

-42-

The terms substantially identical or similar varies with the context as understood by those skilled in the relevant art and generally means at least 40, 60, 80, 90, 95 or 98%.

As used herein, substantially identical to a product means
5 sufficiently similar so that the property is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel
10 electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce
15 substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector (or plasmid) refers to discrete elements that
20 are used to introduce heterologous DNA into cells for either expression or replication thereof. The vectors typically remain episomal, but may be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and
25 mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector

-43-

refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA.

Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein-binding-sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, a composition refers to any mixture of two or more ingredients. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

As used herein, a combination refers to any association between two or more items.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such compositions.

As used herein, a cellular extract refers to a preparation or fraction that is made from a lysed or disrupted cell.

As used herein, the term "subject" refers to animals, plants, insects, and birds and other phyla, genera and species into which nucleic acid molecules may be introduced. Included are higher organisms, such as mammals, fish, insects and birds, including humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, horses, chicken and others.

-44-

As used herein, flow cytometry refers to processes that use a laser based instrument capable of analyzing and sorting out cells and or chromosomes based on size and fluorescence.

As used herein, the abbreviations for any protective groups, amino
5 acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).

B. Recombination systems

10 Site-specific recombination systems typically contain three elements: a pair of DNA sequences (the site-specific recombination sequences) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction between two site-specific recombination sequences.

15 A number of different site-specific recombinase systems are available and/or known to those of skill in the art, including, but not limited to: the Cre/*lox* recombination system using CRE recombinase (see, e.g., SEQ ID Nos. 58 and 59) from the Escherichia coli phage P1 (see, e.g., Sauer (1993) *Methods in Enzymology* 225:890-900; Sauer *et al.* 20 (1990) *The New Biologist* 2:441-449), Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; Odell *et al.* (1990) *Mol Gen Genet.* 223:369-378; Lasko *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:6232-6236; U.S. Patent No. 5,658,772), the FLP/FRT system of yeast using the FLP recombinase (see, SEQ ID Nos. 60 and 61) from the 2 μ episome of 25 *Saccharomyces cerevisiae* (Cox (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:4223; Falco *et al.* (1982) *Cell* 29:573-584; Golic *et al.* (1989) *Cell* 59:499-509; U.S. Patent No. 5,744,336), the resolvases, including Gin recombinase of phage Mu (Maeser *et al.* (1991) *Mol Gen Genet.* 230:170-176; Klippel, A. *et al* (1993) *EMBO J.* 12:1047-1057; see, e.g.,

-45-

SEQ ID Nos. 64-67), Cin, Hin, $\alpha\delta$ Tn3; the Pin recombinase of *E. coli* (see, e.g., SEQ ID Nos. 68 and 69; Enomoto *et al.* (1983) *J. Bacteriol.* 6:663-668), the R/RS system of the pSR1 plasmid of *Zygosaccharomyces rouxii* (Araki *et al.* (1992) *J. Mol. Biol.* 225:25-37; Matsuzaki *et al.* (1990) *J. Bacteriol.* 172: 610-618) and site-specific recombinases from *Kluyveromyces drosophilarius* (Chen *et al.* (1986) *Nucleic Acids Res.* 314:4471-4481) and *Kluyveromyces waltii* (Chen *et al.* (1992) *J. Gen. Microbiol.* 138:337-345). Other systems are known to those of skill in the art (Stark *et al.* *Trends Genet.* 8:432-439; Utatsu *et al.* (1987) *J. Bacteriol.* 169:5537-5545; see, also, U.S. Patent No. 6,171,861).

Members of the highly related family of site-specific recombinases, the resolvase family, such as $\gamma\delta$, Tn3 resolvase, Hin, Gin, and Cin are also available. Members of this family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser *et al.* (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination (see, U.S. Patent No. 6,171,861).

The bacteriophage P1 Cre/lox and the yeast FLP/FRT systems are particularly useful systems for site-specific integration, inversion or excision of heterologous nucleic acid into, and out of, chromosomes, particularly ACes as provided herein. In these systems a recombinase (Cre or FLP) interacts specifically with its respective site-specific recombination sequence (lox or FRT, respectively) to invert or excise the intervening sequences. The sequence for each of these two systems is relatively short (34 bp for lox and 47 bp for FRT).

The FLP/FRT recombinase system has been demonstrated to function efficiently in plant cells (U.S. Patent No. 5,744,386), and, thus, can be used for producing plant artificial chromosome platforms. In

-46-

general, short incomplete FRT sites leads to higher accumulation of excision products than the complete full-length FRT sites. The system catalyzes intra- and intermolecular reactions, and, thus, can be used for DNA excision and integration reactions. The recombination reaction is
5 reversible and this reversibility can compromise the efficiency of the reaction in each direction. Altering the structure of the site-specific recombination sequences is one approach to remedying this situation. The site-specific recombination sequence can be mutated in a manner that the product of the recombination reaction is no longer recognized as
10 a substrate for the reverse reaction, thereby stabilizing the integration or excision event.

In the Cre-lox system, discovered in bacteriophage P1, recombination between loxP sites occurs in the presence of the Cre recombinase (see, *e.g.*, U.S. Patent No. 5,658,772). This system can be
15 used to insert, invert or excise nucleic acid located between two lox sites. Cre can be expressed from a vector. Since the lox site is an asymmetrical nucleotide sequence, lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA
20 segment located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule
25 result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the Cre coding region.

-47-

Any site-specific recombinase system known to those of skill in the art is contemplated for use herein. It is contemplated that one or a plurality of sites that direct the recombination by the recombinase are introduced into an artificial chromosome to produce platform ACes. The resulting platform ACes are introduced into cells with nucleic acid encoding the cognate recombinase, typically on a vector, and nucleic acid encoding heterologous nucleic acid of interest linked to the appropriate recombination site for insertion into the platform ACes. The recombinase-encoding-nucleic acid may be introduced into the cells on the same vector, or a different vector, encoding the heterologous nucleic acid.

An *E. coli* phage lambda integrase system for ACes platform engineering and for artificial chromosome engineering is provided (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181). The phage lambda integrase (Landy, A. (1989) *Annu. Rev. Biochem.* 58:913-94) is adapted herein and the cognate *att* sites are provided. Chromosomes, including ACes, engineered to contain one or a plurality of *att* sites are provided, as are vectors encoding a mutant integrase that functions in the absence other factors. Methods using the modified chromosomes and vectors for introduction of heterologous nucleic acid are also provided.

For purposes herein, one or more of the sites (e.g., a single site or a pair of sites) required for recombination are introduced into an artificial chromosome, such as an ACes chromosome. The enzyme for catalyzing site-directed recombination is introduced with the DNA of interest, or separately, or is engineered onto the artificial chromosome under the control of a regulatable promoter.

As described herein, artificial chromosome platforms containing one or multiple recombination sites are provided. The methods and resulting products are exemplified with the lambda phage Att/Int system, but

-48-

similar methods may be used for production of *ACes* platforms with other recombination systems.

- The *Att/Int* system and vectors provided herein are not only intended for engineering *ACes* platforms, but may be used to engineer an
- 5 *Att/Int* system into any chromosome. Introduction of *att* sites into a chromosome will permit engineering of natural chromosomes, such as by permitting targeted integration genes or regulatory regions, and by controlled excision of selected regions. For example, genes encoding a particular trait may be added to a chromosome, such as plant
- 10 chromosome engineered to contain one or plurality of *att* sites. Such chromosomes may be used for screening DNA to identify genes. Large pieces of DNA can be introduced into cells and the cells screened phenotypically to select those having the desired trait.

C. Platforms

- 15 Provided herein are platform artificial chromosomes (platform *ACes*) containing single or multiple site-specific recombination sites. Chromosome-based platform technology permits efficient and tractable engineering and subsequent expression of multiple gene targets. Methods are provided that use DNA vectors and fragments to create platform
- 20 artificial chromosomes, including animal, particularly mammalian, artificial chromosomes, and plant artificial chromosomes. The artificial chromosomes contain either single or multiple sequence-specific recombination sites suitable for the placement of target gene expression vectors onto the platform chromosome. The engineered chromosome-
- 25 based platform *ACes* technology is applicable for methods, including cellular and transgenic protein production, transgenic plant and animal production and gene therapy. The platform *ACes* are also useful for producing a library of *ACes* comprising random portions of a given genome (e.g., a mammalian, plant or prokaryotic genome) for genomic

-49-

screening; as well as a library of cells comprising different and/or mutually exclusive *ACes* therein.

Exemplary of artificial chromosome platforms are those based on *ACes*. *ACes* artificial chromosomes are non-viral, self-replicating nucleic acid molecules that function as a natural chromosome, having all the elements required for normal chromosomal replication and maintenance within the cell nucleus. *ACes* artificial chromosomes do not rely on integration into the genome of the cell to be effective, and they are not limited by DNA carrying capacity and as such the therapeutic gene(s) of interest, including regulatory sequences, can be engineered into the *ACes*. In addition, *ACes* are stable *in vitro* and *in vivo* and can provide predictable long-term gene expression. Once engineered and delivered to the appropriate cell or embryo, *ACes* work independently alongside host chromosomes, for *ACes* that are predominantly heterochromatin producing only the products (proteins) from the genes it carries. As provided herein *ACes* are modified by introduction of recombination site(s) to provide a platform for ready introduction of heterologous nucleic acid. The *ACes* platforms can be used for production of transgenic animals and plants; as vectors for genetic therapy; for use as protein production systems; for animal models to identify and target new therapeutics; in cell culture for the development and production of therapeutic proteins; and for a variety of other applications.

1. Generation of artificial chromosomes

Artificial chromosomes may be generated by any method known to those of skill in the art. Of particular interest herein are the *ACes* artificial chromosomes, which contain a repeated unit. Methods for production of *ACes* are described in detail in U.S. Patent Nos. 6,025,155 and 6,077,697, which, as with all patents, applications, publications and other disclosure, are incorporated herein in their entirety.

-50-

Generation of *de novo* ACes.

ACes can be generated by cotransfecting exogenous DNA—such as a mammary tissue specific DNA cassette including the gene sequences for a therapeutic protein, with a rDNA fragment and a drug resistance marker gene into the desired eukaryotic cell, such as plant or animal cells, such as murine cells *in vitro*. DNA with a selectable or detectable marker is introduced, and can be allowed to integrate randomly into pericentric heterochromatin or can be targeted to pericentric heterochromatin, such as that in rDNA gene arrays that reside on acrocentric chromosomes, such as the short arms of acrocentric chromosomes. This integration event activates the “megareplicator” sequence and amplifies the pericentric heterochromatin and the exogenous DNA, and duplicates a centromere. Ensuing breakage of this “dicentric” chromosome can result in the production of daughter cells that contain the substantially-original chromosome and the new artificial chromosome. The resulting ACes contain all the essential elements needed for stability and replication in dividing cells—centromere, origins of replications, and telomeres. ACes have been produced that express marker genes (lacZ, green fluorescent protein, neomycin-resistance, puromycin-resistance, hygromycin-resistance) and genes of interest. Isolated ACes, for example, have been successfully transferred intact to rodent, human, and bovine cells by electroporation, sonoporation, microinjection, and transfection with lipids and dendrimers.

To render the creation of ACes with desired genes more tractable and efficient, “platform” ACes (platform-ACes) can be produced that contain defined DNA sequences for enzyme-mediated homologous DNA recombination, such as by Cre or FLP recombinases (Bouhassira *et al.* (1996) *Blood* 88(supplement 1):190a; Bouhassira *et al.* (1997) *Blood*, 90:3332-3344; Siebler *et al.* (1997) *Biochemistry*: 36:1740-1747;

-51-

Siebler *et al.* (1998) *Biochemistry* 37: 6229-6234; and Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834), and as exemplified herein the lambda phage integrase. A *lox* site contains two 13 bp inverted repeats to which Cre-recombinase binds and an intervening 8 bp core region.

- 5 Only pairs of sites having identity in the central 6 bp of the core region are proficient for recombination; sites having non-identical core sequences (heterospecific *lox* sites) do not efficiently recombine with each other (Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300).

10 **Generating acrocentric chromosomes for plant artificial chromosome formation.**

In human and mouse cells *de novo* formation of a satellite DNA based artificial chromosome (SATAC, also referred to as ACes) can occur in an acrocentric chromosome where the short arm contains only pericentric heterochromatin, the rDNA array, and telomere sequences.

- 15 Plant species may not have any acrocentric chromosomes with the same physical structure described, but "megareplicator" DNA sequences reside in the plant rDNA arrays, also known as the nucleolar organizing regions (NOR). A structure like those seen in acrocentric mammalian chromosomes can be generated using site-specific recombination between
20 appropriate arms of plant chromosomes.

Approach

- Qin *et al.* ((1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:1706-1710, 1994) describes crossing two *Nicotiana tabacum* transgenic plants. One plant contains a construct encoding a promoterless hygromycin-resistance
25 gene preceded by a *lox* site (*lox-hpt*), the other plant carries a construct containing a cauliflower mosaic virus 35S promoter linked to a *lox* sequence and the *cre* DNA recombinase coding region (35S-*lox-cre*). The constructs were introduced separately by infecting leaf explants with *agrobacterium tumefaciens* which carries the kanamycin-resistance gene

-52-

(Kan^R). The resultant Kan^R transgenic plants were crossed. Plants that carried the appropriate DNA recombination event were identified by hygromycin-resistance.

5 **Modification of the above for generation of ACes**

The Kan^R cultivars are initially screened, such as by FISH, to identify two sets of candidate transgenic plants. One set has one construct integrated in regions adjacent to the pericentric heterochromatin on the short arm of any chromosome. The second set of candidate plants
10 has the other construct integrated in the NOR region of appropriate chromosomes. To obtain reciprocal translocation both sites must be in the same orientation. Therefore a series of crosses are required, Kan^R plants generated, and FISH analyses performed to identify the appropriate "acrocentric" plant chromosome for *de novo* plant ACes formation.

15 **2. Bacteriophage lambda integrase-based site-specific recombination system**

An integral part of the platform technology includes a site-specific recombination system that allows the placement of selected gene targets or genomic fragments onto the platform chromosomes. Any such system
20 may be used. In particular, a method is provided for insertion of additional DNA fragments into the platform chromosome residing in the cell via sequence-specific recombination using the recombinase activity of the bacteriophage lambda integrase. The lambda integrase system is exemplary of the recombination systems contemplated for ACes. Any
25 known recombination system, including any described herein, particularly any that operates without the need for additional factors or that, by virtue of mutation, does not require additional factors, is contemplated.

-53-

As noted the lambda integrase system provided herein can be used with natural chromosomes and artificial chromosomes in addition to ACes. Single or a plurality of recombination sites, which may be the same or different, are introduced into artificial chromosomes to produce
5 artificial chromosome platforms.

3. Creation of bacteriophage lambda integrase site-specific recombination system

The lambda phage-encoded integrase (designated Int) is a prototypical member of the integrase family. Int effects integration and
10 excision of the phage in and out of the *E. coli* genome via recombination between pairs of attachment sites designated *attB/attP* and *attL/attR*. Each *att* site contains two inverted 9 base pair core Int binding sites and a 7 base pair overlap region that is identical in wild-type *att* sites. Each site, except for *attB* contains additional Int binding sites. In flanking
15 regions, there are recognition sequences for accessory DNA binding proteins, such as integration host factor (IHF), factor for inversion stimulation (FIS) and the phage encoded excision protein (XIS). Except for *attB*, Int is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins and negative DNA supercoiling, binds
20 simultaneously to core and arm sites within the same *att* site.

Int, like Cre and FLP, executes an ordered sequential pair of strand exchanges during integrative and excisive recombination. The natural pairs of target sequences for Int, *attB* and *attP* or *attL* and *attR* are located on the same or different DNA molecules resulting in intra or
25 intermolecular recombination, respectively. For example, intramolecular recombination occurs between inversely oriented *attB* and *attP*, or between *attL* and *attR* sequences, respectively, leading to inversion of the intervening DNA segment.

-54-

Like the recombinase systems, such as Cre and FLP, Int directs site-specific recombination. Unlike the other systems, such as Cre and FLP, Int generally requires additional protein factors for integrative and excisive recombination and negative supercoiling for integrative recombination.

5 Hence, the Int system had not been used in eukaryotic targeting systems.

Mutant Int proteins, designated Int-h (E174K) and a derivative thereof Int-h/218(E174K/E218K) do not require accessory proteins to perform intramolecular integrative and excisive recombination in co-transfection assays in human cells (Lorbach *et al.* (2000) *J Mol. Biol.* 10 296:1175-1181); wild-type Int does not catalyze intramolecular recombination in human cells harboring target sites *attB* and *attP*. Hence it had been demonstrated that mutant Int can catalyze factor-independent recombination events in human cells.

There has been no demonstration by others that this system can be 15 used for engineering of eukaryotic genomes or chromosomes. Provided herein are chromosomes, including artificial chromosomes, such as but not limited to ACes that contain *att* sites (e.g., platform ACes), and the use of such chromosomes for targeted integration of heterologous DNA into such chromosomes in eukaryotic cells, including animal, such as 20 rodent and human, and plant cells. Mutant Int provided herein is shown to effect site-directed recombination between sites in artificial chromosomes and vectors containing cognate sites.

An additional component of the chromosome-based platform technology is the site-specific integration of target DNA sequences onto 25 the platform. For this the native bacteriophage lambda integrase has been modified to carry out this sequence specific DNA recombination event in eukaryotic cells. The bacteriophage lambda integrase and its cognate DNA substrate *att* is a member of the site-specific recombinase family that also includes the bacteriophage P1 Cre/lox system as well as

-55-

the *Saccharomyces cerevisiae* 2 micron based FLP/FRT system (see, e.g., Landy (1989) *Ann. Rev. Biochem* 58:913-949; Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402; Broach *et al.* (1982) *Cell* 29:227-234).

- 5 By combining DNA endonuclease and DNA ligase activity these recombinases recognize and catalyze DNA exchanges between sequences flanking the recognition site. During the integration of lambda genome into the *E. coli* (lambda recombination) genome, the phage integrase (INT) in association with accessory proteins catalyzes the DNA exchange
- 10 between the attP site of the phage genome and the attB site of the bacterial genome resulting in the formation of attL and attR sites (Figure 6). The engineered bacteriophage lambda integrase has been produced herein to carry out an intermolecular DNA recombination event between
- 15 attB site) and the chromosome-based platform carrying the lambda attP sequence independent of lambda bacteriophage or bacterial accessory proteins.

- In contrast to the bi-directional Cre/lox and FLP/FRT system, the engineered lambda recombination system derived for chromosome-based
- 20 platform technology is advantageously unidirectional because accessory proteins, which are absent, are required for excision of integrated nucleic acid upon further exposure to the lambda Int recombinase.

4. Creation of platform chromosome containing single or multiple sequence-specific recombination sites

25 **a. Multiple sites**

For the creation of a platform chromosome containing multiple, sequence-specific recombination sites, artificial chromosomes are produced as depicted in Figure 5 and Example 3. As discussed above, artificial chromosomes can be produced using any suitable methodology,

-56-

including those described in U.S. Patent Nos. 5,288,625; 5,712,134; 5,891,691; 6,025,155. Briefly, to prepare artificial chromosomes containing multiple recombination (e.g., integration) sites, nucleic acid (either in the form a one or more plasmids, such as the plasmid pSV40193attPsensePUR set forth in Example 3) is targeted into an amplifiable region of a chromosome, such as the pericentric region of a chromosome. Among such regions are the rDNA gene loci in acrocentric mammalian chromosomes. Hence, targeting nucleic acid for integration into the rDNA region of mammalian acrocentric chromosomes can include the mouse rDNA fragments (for targeting into rodent cell lines) or large human rDNA regions on BAC/PAC vectors (or subclones thereof in standard vectors) for targeting into human acrocentric chromosomes, such as for human gene therapy applications. The targeting nucleic acid generally includes a detectable or selectable marker, such as antibiotic resistance, such as puromycin and hygromycin, a recombination site (such as attP, attB, attL, attR or the like), and/or human selectable markers as required for gene therapy applications. Cells are grown under conditions that result in amplification and ultimately production of ACes artificial chromosomes having multiple recombination (e.g., integration) sites therein. ACes having the desired size are selected for further engineering.

b. Creation of platform chromosome containing a single sequence-specific recombination site

In this method a mammalian platform artificial chromosome is generated containing a single sequence-specific recombination site. In the Example below, this approach is demonstrated using a puromycin resistance marker for selection and a mouse rDNA fragment for targeting into the rDNA locus on mouse acrocentric chromosomes. Other selection markers and targeting DNA sequences as desired and known to those of

-57-

skill in the art can be used. Additional resistance markers include genes conferring resistance to the antibiotics neomycin, blasticidin, hygromycin and zeocin. For applications, such as gene therapy in which potentially immunogenic responses are to be avoided, host, such as human, derived

5 selectable markers or markers detectable with monoclonal antibodies (MAb) followed by fluorescent activated cell sorting (FACS) can be used. Examples in this class include, but are not limited to: human nerve growth factor receptor (detection with MAb); truncated human growth factor receptor (detection with MAb); mutant human dihydrofolate reductase

10 (DHFR; detectable using a fluorescent methotrexate substrate); secreted alkaline phosphatase (SEAP; detectable with fluorescent substrate); thymidylate synthase (TS; confers resistance to fluorodeoxyuridine); human CAD gene (confers resistance to N-phosphonacetyl-L-aspartate (PALA)).

15 To construct a platform artificial chromosome with a single site, an *ACes* artificial chromosome (or other artificial chromosome of interest) can be produced containing a selectable marker. A single sequence specific recombination site is targeted onto *ACes* via homologous recombination. For this, DNA sequences containing the site-specific

20 recombination sequence are flanked with DNA sequences homologous to a selected sequence in the chromosome. For example, when using a chromosome containing rDNA or satellite DNA, such DNA can be used as homologous sequences to target the site-specific recombination sequence onto the chromosome. A vector is designed to have these homologous

25 sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the chromosome, the DNA is transfected into cells harboring the chromosome. After transfection and integration of the site-specific cassette, homologous recombination events onto the platform

-58-

chromosome are subcloned and identified, for example by screening single cell subclones via expression of resistance or a fluorescent marker and PCR analysis. In one embodiment, a platform artificial chromosome, such as a platform ACes, that contains a single copy of the recombination site is selected. Examples 2B and 2D exemplify the process, and Figure 3 provides a diagram depicting one method for the creation of a platform mammalian chromosome containing a single sequence-specific recombination site.

5. **Lambda integrase mediated recombination of target gene expression vector onto platform chromosome**

The third component of the chromosome-based platform technology involves the use of target gene expression vectors carrying, for example, genes for gene therapy, genes for transgenic animal or plant production, and those required for cellular protein production of interest. Using lambda integrase mediated site-specific recombination, or any other recombinase-mediated site-specific recombination, the target gene expression vectors are introduced onto the selected chromosome platform. The use of target gene expression vector permits use of the *de novo* generated chromosome-based platforms for a wide range of gene targets. Furthermore, chromosome platforms containing multiple *attP* sites provides the opportunity to incorporate multiple gene targets onto a single platform, thereby providing for expression of multiple gene targets, including the expression of cellular and genetic regulatory genes and the expression of all or parts of metabolic pathways. In addition to expressing small target genes, such as cDNA and hybrid cDNA/artificial intron constructs, the chromosome-based platform can be used for engineering and expressing large genomic fragments carrying target genes along with its endogenous genomic promoter sequences. This is of importance, for example, where the therapy requires precise cell specific

-59-

expression and in instances where expression is best achieved from genomic clones rather than cDNA clones. Figure 9 provides a diagram summarizing one embodiment of the chromosome-based technology.

A feature of the target gene expression vector that is of interest to
5 include is a promoterless marker gene, which as exemplified (see, Figure 9) contains an upstream *attB* site (marker 2 on Figure 9). The nucleic acid encoding the marker is not expressed unless it is placed downstream from a promoter sequence. Using the recombinase technology provided herein, such as the lambda integrase technology (λ INT_{E174R} on figure 8)
10 provided herein, site-specific recombination between the *attB* site on the vector and the promoter-*attP* site (in the "sense" orientation) on the chromosome-based platform results in the expression of marker 2 on the target gene expression vector, thereby providing a positive selection for the lambda INT mediated site-specific recombination event. Site-specific
15 recombination events on the chromosome-based platform versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. Examples of suitable marker 2 genes, include, but are not limited to, genes that confer resistance to toxic compounds or antibiotics,
20 fluorescence activated cell sorting (FACS) sortable cell surface markers and various fluorescent markers. Examples of these genes include, but are not limited to, human L26a^R (human homolog of *Saccharomyces cerevisiae* CYH^B gene), neomycin, puromycin, blasticidin, CD24 (see, e.g., US Patents 5,804,177 and 6,074,836), truncated CD4, truncated low
25 affinity nerve growth factor receptor (LNGFR), truncated LDL receptor, truncated human growth hormone receptor, GFP, RFP, BFP.

The target gene expression vectors contain a gene (target gene) for expression from the chromosome platform. The target gene can be expressed using various constitutive or regulated promoter systems

-60-

- across various mammalian species. For the expression of multiple target genes within the same target gene expression vector, the expression of the multiple targets can be coordinately regulated via viral-based or human internal ribosome entry site (IRES) elements (see, *e.g.*, Jackson *et al.* (1990) *Trends Biochem Sci.* 15: 477-83; Oumard *et al.* (2000) *Mol. Cell. Biol.* 20: 2755-2759). Furthermore, using IRES type elements linked to a downstream fluorescent marker, *e.g.*, green, red or blue fluorescent proteins (GFP, RFP, BFP) allows for the identification of high expressing clones from the integrated target gene expression vector.
- 10 In certain embodiments described herein, the promoterless marker can be transcriptionally downstream of the heterologous nucleic acid, wherein the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.
- 15 In addition, the selectable marker and the heterologous nucleic acid can be transcriptionally linked by the presence of a IRES between them. As set forth herein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.
- 20 Expression from the target gene expression vector integrated onto the chromosome-based platform can be further enhanced using genomic insulator/boundary elements. The incorporation of insulator sequences into the target gene expression vector helps define boundaries in chromatin structure and thus minimizes influence of chromatin position effects/gene silencing on the expression of the target gene (Bell *et al.* (1999) *Current Opinion in Genetics and Development* 9:191-198; Emery *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:9150-9155). Examples of insulator elements that can be included onto target gene expression vector in order to optimize expression include, but are not limited to:
- 25

-61-

- 1) chicken β -globin HS4 element (Prioleau *et al.* (1999) *EMBO J* 18: 4035-4048);
- 2) matrix attachment regions (MAR; see, *e.g.*, Ramakrishnan *et al.* (2000) *Mol Cell. Biol.* 20:868-877);
- 5 3) scaffold attachment regions (SAR; see, *e.g.*, Auten *et al.* (1999) *Human Gene Therapy* 10:1389-1399); and
- 4) universal chromatin opening elements (UCOE; WO/0005393 and WO/0224930)

The copy number of the target gene can be controlled by sequentially adding multiple target gene expression vectors containing the target gene onto multiple integration sites on the chromosome platform. Likewise, the copy number of the target gene can be controlled within an individual target gene expression vector by the addition of DNA sequences that promote gene amplification. For example, gene amplification can be induced utilizing the dihydrofolate reductase (DHFR) minigene with subsequent selection with methotrexate (see, *e.g.*, Schimke (1984) *Cell* 37:705-713) or amplification promoting sequences from the rDNA locus (see, *e.g.*, Wegner *et al.* (1989) *Nucl. Acids Res.* 17: 9909-9932).

20 6. Platforms with other recombinase system sites

A "double *lox*" targeting strategy mediated by Cre-recombinase (Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834) can be used. This strategy employs a pair of heterospecific *lox* sites—*loxA* and *loxB*, which differ by one nucleotide in the 8 bp spacer region. Both sites are engineered into the artificial chromosome and also onto the targeting DNA vector. This allows for a direct site-specific insertion of a commercially relevant gene or genes by a Cre-catalyzed double crossover event. In essence a platform ACes is engineered with a hygromycin-resistance gene flanked by the double *lox* sites generating *lox-ACes*, which is maintained

-62-

in the thymidine kinase deficient cell, LMtk(-). The gene of interest, for example, for testing purposes, the green fluorescence protein gene, GFP and a HSV thymidine kinase gene (tk) marker, are engineered between the appropriate *lox* sites of the targeting vector. The vector DNA is

5 cotransfected with plasmid pBS185 (Life Technologies) encoding the *Cre* recombinase gene into mammalian cells maintaining the dual-*lox* artificial chromosome. Transient expression of the *Cre* recombinase catalyzes the site-specific insertion of the gene and the tk-gene onto the artificial chromosome. The transfected cells are grown in HAT medium that

10 selects for only those cells that have integrated and expressed the thymidine kinase gene. The HAT^R colonies are screened by PCR analyses to identify artificial chromosomes with the desired insertion.

To generate the *lox-ACes*, Lambda-Hyg^R-*lox* DNA is transfected into the LMtk(-) cell line harboring the precursor *ACes*. Hygromycin-

15 resistant colonies are analyzed by FISH and Southern blotting for the presence of a single copy insert on the *ACes*.

To demonstrate the gene replacement technology, cell lines containing candidate *lox-ACes* are cotransfected with pTK-GFP-*lox* and pBS185 (encoding the *Cre* recombinase gene) DNA. After transfection,

20 transient expression of plasmid pBS185 will provide sufficient burst of *Cre* recombinase activity to catalyze DNA recombination at the *lox* sites. Thus, a double crossover event between the *ACes* target and the exogenous targeting plasmid carrying the *loxA* and *loxB* permits the simple replacement of the hygromycin-resistance gene on the *lox-ACes*

25 for the tk-GFP cassette from the targeting plasmid, with no integration of vector DNA. Transfected cells are grown in HAT-media to select for tk-expression. Correct targeting will result in the generation of HAT^R, hygromycin sensitive, and green fluorescent cells. The desired integration event is verified by Southern and PCR analyses. Specific PCR primer sets

-63-

are used to amplify DNA sequences flanking the individual *loxA* and *loxB* sites on the *lox-ACes* before and after homologous recombination.

D. Exemplary applications of the Platform ACes

Platform *ACes* are applicable and tractable for different/optimized
5 cell lines. Those that include a fluorescent marker, for example, can be purified and isolated using fluorescent activated cell sorting (FACS), and subsequently delivered to a target cell. Those with selectable markers provide for efficient selection and provide a growth advantage. Platform *ACes* allow multiple payload delivery of donor target vectors via a
10 positive-selection site-specific, recombination system, and they allow for the inclusion of additional genetic factors that improve protein production and protein quality.

The construction and use of the platform *ACes* as provided for each application may be similarly applied to other applications. Particular
15 descriptions are for exemplification.

1. Cellular Protein Production Platform ACes (CPP ACes)

As described herein, *ACes* can be produced from acrocentric chromosomes in rodent (mouse, hamster) cell lines via megareplicator induced amplification of heterochromatin/rDNA sequences. Such *ACes*
20 are ideal for cellular protein production as well as other applications described herein and known to those of skill in the art. *ACes* platforms that contain a plurality of recombination sites are particularly suitable for engineering as cellular protein production systems.

In one embodiment, CPP *ACes* involve a two-component system:
25 the platform chromosome containing multiple engineering sites and the donor target vector containing a platform-specific recombination site with designed expression cassettes (see Figure 9).

The platform *ACes* can be produced from any artificial chromosome, particularly the amplification-based artificial chromosomes.

-64-

For exemplification, they are produced from rodent artificial chromosomes produced from acrocentric chromosomes using the technology of U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183, in which nucleic acid is targeted to the pericentric heterochromatic, and, particularly into rDNA to initiate the replication event(s). The ACes can be produced directly in the chosen cellular protein production cell lines, such as, but not limited to, CHO cells, hybridomas, plant cells, plant tissues, plant protoplasts, stem cells and plant calli.

10 a. **Platform Construction**

In the exemplary embodiment, the initial *de novo* platform construction requires co-transfecting with excess targeting DNA, such as, rDNA or lambda DNA without an *attP* region, and an engineered selectable marker. The engineered selectable marker should contain
15 promoter, generally a constitutive promoter, such as human, viral, i.e., adenovirus or SV40 promoter, including the human ferritin heavy chain promoter (SEQ ID NO:128), SV40 and EF1 α promoters, to control expression of a marker gene that provides a selective growth advantage to the cell. An example of such a marker gene is the *E. coli hisD* gene
20 (encoding histidinol dehydrogenase) which is homologous and analogous to the *S. typhimurium hisD* a dominant marker selection system for mammalian cells previously described (see, Hartman *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051). Since histidine is an essential amino acid in mammals and a nutritional requirement in cell culture, the *E.*
25 *coli hisD* gene can be used to select for histidine prototrophy in defined media. Furthermore more stringent selection can be placed on the cells by including histinol in the medium. Histidinol is itself permeable and toxic to cells. The *hisD* provides a means of detoxification.

-65-

Placed between the promoter and the marker gene is the bacteriophage lambda *attP* site to use the bacteriophage lambda integrase dependent site-specific recombination system (described herein). The insertion of an *attP* site downstream of a promoter element provide
5 forward selection of site-specific recombination events onto the platform *ACes*.

b. Donor Target Vector Construction

A second component of the CPP platform *ACes* system involves the construction of donor target vectors containing a gene product(s) of
10 interest for the CPP platform *ACes*. Individual donor target vectors can be designed for each gene product to be expressed thus enabling maximum usage of a *de novo* constructed platform *ACes*, so that one or a few CPP platform *ACes* will be required for many gene targets.

A key feature of the donor vector target is the *promoterless* marker
15 gene containing an upstream *attB* site (marker 2 on figure 9). Normally the marker would not be expressed unless it is placed downstream of a promoter sequence. As discussed above, using the lambda integrase technology (λ INT_{E174R} on Figure 8 and Figure 9), site-specific recombination between the *attB* site on the vector and the promoter-*attP*
20 site on the CPP platform *ACes* result in the expression of the donor target vector marker providing positive selection for the site-specific event. Site-specific recombination events on the CPP *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.
25 In addition, since the lambda integrase reaction is unidirectional, i.e. excision reaction is not possible, a number of unique targets can be loaded onto the CPP platform *ACes* limited only by the number of markers available.

-66-

Additional features of the donor target vector include gene target expression cassettes flanked by either chromatin insulator regions, matrix attachment regions (MAR) or scaffold attachment regions (SAR). The use of these regions will provide a more "open" chromatin environment for gene expression and help alleviate silencing. An example of such a cassette for expressing a monoclonal antibody is described. For this purpose, a strong constitutive promoter, e.g. chicken β -actin or RNA Poll, is used to drive the expression of the heavy and light chain open reading frames. The heavy and light chain sequences flank a nonattenuated human IRES (IRES_H; from the 5'UTR of NRF1 gene; see Oumard et al., 2000, *Mol. and Cell Biol.*, 20(8):2755-2759) element thereby coordinating transcription of both heavy and light chain sequence. Distal to the light chain open reading frame resides an additional viral encoded IRES (IRES_V modified ECMV internal ribosomal entry site (IRES)) element attenuating the expression of the fluorescent marker gene hrGFP from *Renilla* (Stratagene). By linking the hrGFP with an attenuated IRES, the heavy and light chains along with the hrGFP are monocistronic. Thus, the identification of hrGFP fluorescing cells will provide a means to detect protein producing cells. In addition, high producing cell lines can be identified and isolated by FACS thereby decreasing the time frame in finding high expressers. Functional monoclonal antibody will be confirmed by ELISA.

c. Additional components in cellular protein production platform ACes (CPP ACes)

In addition to the aforementioned CPP ACes system, other genetic factors can be included to enhance the yield and quality of the expressed protein. Again to provide maximum flexibility, these additional factors can be inserted onto the CPP platform ACes by λ INTE174R dependent site-specific recombination. Other factors that could be used with a CPP

-67-

Platform *ACes* include for example, adenovirus E1a transactivation system which upregulates both cellular and viral promoters (see, e.g., Svensson and Akusjarvi (1984) EMBO 3:789-794; and US patents 5,866,359; 4,775,630 and 4,920,211).

5 **d. Targets for CHO-*ACes* engineering to enhance cell growth, such as CHO cell growth and protein production/ quality**

 If adding these additional factors onto the CPP *ACes* is not prudent or desired, the host cell, CHO cells, can be engineered to express these
10 factors (see, below, targets for CHO-*ACes* engineering to enhance CHO cell growth and protein production/quality). Additional factors to consider including are addition of insulin or IGF-1 to sustain viability;
human sialyltransferases or related factors to produce more human-like glycoproteins; expression of factors to decrease ammonium accumulation
15 during cell growth; expression of factors to inhibit apoptosis; expression of factors to improve protein secretion and protein folding; and expression of factors to permit serum-free transfection and selection.

1) Addition of insulin or IGF-1 to sustain viability

20 Stimulatory factors and/or their receptors are expressed to set up an autocrine loop, to improve cell growth, such as CHO cell growth. Two exemplary candidates are insulin and IGF-1 (see, Biotechnol Prog 2000 Sep;16(5):693-7). Insulin is the most commonly used growth factor for sustaining cell growth and viability in serum-free Chinese hamster ovary
25 (CHO) cell cultures. Insulin and IGF-1 analog (LongR(3) serve as growth and viability factors for CHO cells.

 CHO cells were modified to produce higher levels of essential nutrients and factors. A serum-free (SF) medium for dihydrofolate reductase-deficient Chinese hamster ovary cells (DG44 cells) was
30 prepared. Chinese hamster ovary cells (DG44 cells), which are normally

-68-

maintained in 10% serum medium, were gradually weaned to 0.5% serum medium to increase the probability of successful growth in SF medium (see, Kim *et al.* (199) *In Vitro Cell Dev Biol Anim* 35(4):178-82). A SF medium (SF-DG44) was formulated by supplementing the basal medium with these components; basal medium was prepared by supplementing Dulbecco's modified Eagle's medium and Ham's nutrient mixture F12 with hypoxanthine (10 mg/l) and thymidine (10 mg/l). Development of a SF medium for DG44 cells was facilitated using a Plackett-Burman design technique and weaning of cells.

10

2) Human sialyltransferases or related factors to produce more human-like glycoproteins

CHO cells have been modified by increasing their ability to process protein via addition of complex carbohydrates. This has been achieved by overexpression of relevant processing enzymes, or in some cases, reducing expression of relevant enzymes (see, Bragonzi *et al.* (2000) *Biochim Biophys Acta* 1474(3):273-282; see, also Weikert *et al.* (1999) *Nature biotech.* 17:1116-11121; Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). A CHO cell line expressing alpha2,6-sialyltransferase was developed for the production of human-like sialylated recombinant glycoproteins. The sialylation defect of CHO cells can be corrected by transfecting the alpha2,6-sialyltransferase (alpha2,6-ST) cDNA into the cells. Glycoproteins produced by such CHO cells display alpha2,6-and alpha2,3-linked terminal sialic acid residues, similar to human glycoproteins.

As another example for improving the production of human-like sialylated recombinant glycoproteins, a CHO cell line has been developed that constitutively expresses sialidase antisense RNA (see, Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). Several antisense expression

30

-69-

vectors were prepared using different regions of the sialidase gene. Co-transfection of the antisense constructs with a vector conferring puromycin resistance gave rise to over 40 puromycin resistant clones that were screened for sialidase activity. A 5' 474 bp coding segment of the sialidase cDNA, in the inverted orientation in an SV 40-based expression vector, gave maximal reduction of the sialidase activity to about 40% wild-type values.

Oligosaccharide biosynthesis pathways in mammalian cells have been engineered for generation of recombinant glycoproteins (see, *e.g.*, Sburlati (1998) *Biotechnol Prog* 14(2):189-92), which describes a Chinese hamster ovary (CHO) cell line capable of producing bisected oligosaccharides on glycoproteins. This cell line was created by overexpression of a recombinant N-acetylglucosaminyltransferase III (GnT-III) (see, also, Prati *et al.* (1998) *Biotechnol Bioeng* 59(4):445-50, which describes antisense strategies for glycosylation engineering of CHO cells).

3) Expression of factors to decrease ammonium accumulation during cell growth

Excess ammonium, which is a by-product of CHO cell metabolism can have detrimental effects on cell growth and protein quality (see, Yang *et al.* (2000) *Biotechnol Bioeng* 68(4):370-80). To solve this problem ammonium levels were modified by overexpressing carbamoyl phosphate synthetase I and ornithine transcarbamoylase or glutamine synthetase in CHO cells. Such modification resulted in reduced ammonium levels observed and an increase in the growth rate (see Kim *et al.* (2000) *J Biotechnol* 81(2-3):129-40; and Enosawa *et al.* (1997) *Cell Transplant* 6(5):537-40).

4) Expression of factors to improve protein secretion and protein folding

-70-

Overexpression of relevant enzymes can be engineered into the ACes to improve protein secretion and folding.

5) Expression of factors to permit serum-free transfection and selection

5 It is advantageous to have the ability to convert CHO cells in suspension growing in serum free medium to adherence with out having to resort to serum addition. Laminin or fibronectin addition is sufficient to make cells adherent (see, *e.g.*, Zaworski *et al.* (1993) *Biotechniques* 15(5):863-6) so that expressing either of these genes in CHO cells under
10 an inducible promoter should allow for reversible shift to adherence without requiring serum addition.

2. Platform ACes and Gene Therapy

The platform ACes provided herein are contemplated for use in mammalian gene therapy, particularly human gene therapy. Human ACes
15 can be derived from human acrocentric chromosomes from human host cells, in which the amplified sequences are heterochromatic and/or human rDNA. Different platform ACes applicable for different tissue cell types are provided. The ACes for gene therapy can contain a single copy of a therapeutic gene inserted into a defined location on platform ACes.

20 Therapeutic genes include genomic clones, cDNA, hybrid genes and other combinations of sequences. Preferred selectable markers are those from the mammalian host, such as human derived factors so that they are non-immunogenic, non-toxic and allow for efficient selection, such as by FACS and/or drug resistance.

25 Platform ACes, useful for gene therapy and other applications, as noted herein, can be generated by megareplicator dependent amplification, such as by the methods in U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183. In one embodiment, human ACes are produced using

-71-

human rDNA constructs that target rDNA arrays on human acrocentric chromosomes and induce the megareplicator in human cells, particularly in primary cell lines (with sufficient number of doublings to form the ACes) or stem cells (such as hematopoietic stem cells, mesenchymal stem cells, adult stem cells or embryonic stem cells) to avoid the introduction of potentially harmful rearranged DNA sequences present in many transformed cell lines. Megareplicator induced ACes formation can result in multiple copies of targeting DNA/selectable markers in each amplification block on both chromosomal arms of the platform ACes.

10 In view of the considerations regarding immunogenicity and toxicity, the production of human platform ACes for gene therapy applications employs a two component system analogous to the platform ACes designed for cellular protein production (CPP platform ACes). The system includes a platform chromosome of entirely human DNA origin
15 containing multiple engineering sites and a gene target vector carrying the therapeutic gene of interest.

a. Platform Construction

The initial *de novo* construction of the platform chromosome employs the co-transfection of excess targeting DNA and a selectable
20 marker. In one embodiment, the DNA is targeted to the rDNA arrays on the human acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22). For example, two large human rDNA containing PAC clones 18714 and 18720 and the human PAC clone 558F8 are used for targeting (Genome Research (ML) now Incyte, BACPAC Resources, 747 52nd
25 Street, Oakland CA). The mouse rDNA clone pFK161 (SEQ ID NO: 118), which was used to make the human SATAC from the 94-3 hamster/human hybrid cell line (see, *e.g.*, published International PCT application No. WO 97/40183 and Csonka, *et al*, *Journal of Cell Science*

-72-

113:3207-32161 and Example 1 for a description of pFK161) can also be used.

For animal applications, selectable markers should be non-immunogenic in the animal, such as a human, and include, but are not limited to: human nerve growth factor receptor (detected with a MAb, such as described in US patent 6,365,373); truncated human growth factor receptor (detected with MAb), mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine); human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34+ cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS); human CD25 (IL-2 α ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); and Cytidine deaminase (CD; selectable by Ara-C).

Since megareplicator induced amplification generates multiple copies of the selectable marker, a second consideration for the selection of the human marker is the resulting dose of the expressed marker after ACes formation. High level of expression of certain markers may be detrimental to the cell and/or result in autoimmunity. One method to decrease the dose of the marker protein is by shortening its half-life, such as via the fusion of the well-conserved human ubiquitin tag (a 76 amino acid sequence) thus leading to increased turnover of the selectable marker. This has been used successfully for a number of reporter

-73-

systems including DHFR (see, *e.g.*, Stack *et al.* (2000) *Nature Biotechnology* 18:1298-1302 and references cited therein).

Using the ubiquitin tagged protein, a human selectable marker system analogous to the CPP ACes described herein is constructed.

- 5 Briefly, a tagged selectable marker, such as for example one of those described herein, is cloned downstream of an *attP* site and expressed from a human promoter. Exemplary promoters contemplated for use herein include, but are not limited to, the human ferritin heavy chain promoter (SEQ ID NO:128); RNA Poll; EF1 α ; TR; glyceraldehyde-3-
10 phosphate dehydrogenase core promoter (GAP); a GAP core promoter including a proximal insulin inducible element the intervening GAP sequence; phosphofructokinase promoter; and phosphoglycerate kinase promoter. Also contemplated herein is an aldolase A promoter H1 & H2 (representing closely spaced transcriptional start sites) along with the
15 proximal H enhancer. There are 4 promoters (*e.g.*, transcriptional start sites) for this gene, each having different regulatory and tissue activity. The H (most proximal 2) promoters are ubiquitously expressed off the H enhancer. This resulting marker can then be co-transfected along with excess human rDNA targeting sequence into the host cells. An important
20 criteria for the selection of the

- recipient cells is sufficient number of cell doublings for the formation and detection of ACes. Accordingly, the co-transfections should be attempted in human primary cells that can be cultured for long periods of
25 time, such as for example, stem cells (*e.g.*, hematopoietic, mesenchymal, adult or embryonic stem cells), or the like. Additional cell types, include, but are not limited to: single gene transfected cells exhibiting increased life-span; over-expressing c-myc cells, *e.g.* MSU1.1 (Morgan *et al.*, 1991, *Exp. Cell Res.*, Nov;197(1):125-136); over-expressing telomerase lines,

-74-

such as TERT cells; SV40 large T-antigen transfected lines; tumor cell lines, such as HT1080; and hybrid human cell lines, such as the 94-3 hamster/human hybrid cell line.

b. Gene Target Vector

5 The second component of the GT platform *ACes* (GT *ACes*) system involves the use of engineered target vectors carrying the therapeutic gene of interest. These are introduced onto the GT platform *ACes* via site-specific recombination. As with the CPP *ACes*, the use of engineered target vectors maximizes the use of the *de novo* generated GT platform
10 *ACes* for most gene targets. Furthermore, using lambda integrase technology, GT platform *ACes* containing multiple *attP* sites permits the opportunity to incorporate multiple therapeutic targets onto a single platform. This could be of value in cases where a defined therapy requires multiple gene targets, a single therapeutic target requires an
15 additional gene regulatory factor or a GT *ACes* requires a "kill" switch.

Similar to the CPP *ACes*, a feature of the gene target vector is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on Figure 9). Normally, the marker (in this case, a cell surface antigen that can be sorted by FACS would be ideal) would not be expressed unless it
20 is placed downstream of a promoter sequence. Using the lambda integrase technology (λ INT_{E174R} on figure 9), site-specific recombination between the *attB* site on the vector and the promoter- *attP* site on the GT platform *ACes* results in the expression of marker#2 on the gene target vector, i.e. positive selection for the site-specific event. Site-specific
25 recombination events on the GT *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.

For expression of the therapeutic gene, human specific promoters, such as a ferritin heavy chain promoter (SEQ ID NO:128); EF1 α or RNA

-75-

Poll, are used. These promoters are for high level expression of a cDNA encoded therapeutic protein. In addition to expressing cDNA (or even hybrid cDNA/artificial intron constructs), the GT platform ACes are used for engineering and expressing large genomic fragments carrying
5 therapeutic genes of interest expressed from native promoter sequences. This is of importance in situations where the therapy requires precise cell specific expression or in instances where expression is best achieved from genomic clones versus cDNA.

10 **3. Selectable markers for use, for example, in Gene Therapy (GT)**

The following are selectable markers that can be incorporated into human ACes and used for selection.

15 **Dual Resistance to 4-Hydroperoxycyclophosphamide and Methotrexate by Retroviral Transfer of the Human Aldehyde Dehydrogenase Class 1 Gene and a Mutated Dihydrofolate Reductase Gene**

The genetic transfer of drug resistance to hematopoietic cells is one approach to overcoming myelosuppression caused by high-dose chemotherapy. Because cyclophosphamide (CTX) and methotrexate
20 (MTX) are commonly used non-cross-resistant drugs, generation of dual drug resistance in hematopoietic cells that allows dose intensification may increase anti-tumor effects and circumvent the emergence of drug-resistant tumors, a retroviral vector containing a human cytosolic ALDH-1-encoding DNA clone and a human doubly mutated DHFR-encoding
25 clone (Phe22/Ser31; termed F/S in the description of constructs) to generate increased resistance to CTX and MTX were constructed (Takebe *et al.* (2001) *Mol Ther* 3(1):88-96). This construct may be useful for protecting patients from high-dose CTX- and MTX-induced myelosuppression. ACes can be similarly constructed.

-76-

5 **Multiple mechanisms of N-phosphonacetyl-L-aspartate resistance in human cell lines: carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase gene amplification is frequent only when chromosome 2 is rearranged**

 Rodent cells resistant to N-phosphonacetyl-L-aspartate (PALA) invariably contain amplified carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase (CAD) genes, usually in widely spaced tandem arrays present as extensions of the same chromosome arm that carries a single copy of CAD in normal cells (Smith *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1816-21). In contrast, amplification of CAD is very infrequent in several human tumor cell lines. Cell lines with minimal chromosomal rearrangement and with unrearranged copies of chromosome 2 rarely develop intrachromosomal amplifications of CAD.

15 These cells frequently become resistant to PALA through a mechanism that increases the aspartate transcarbamylase activity with no increase in CAD copy number, or they obtain one extra copy of CAD by forming an isochromosome 2p or by retaining an extra copy of chromosome 2. In cells with multiple chromosomal aberrations and rearranged copies of

20 chromosome 2, amplification of CAD as tandem arrays from rearranged chromosomes is the most frequent mechanism of PALA resistance. All of these different mechanisms of PALA resistance are blocked in normal human fibroblasts. Thus, ACes with multiple copies of the CAD gene would provide PALA resistance.

25 **Retroviral coexpression of thymidylate synthase and dihydrofolate reductase confers fluoropyrimidine and antifolate resistance**

 Retroviral gene transfer of dominant selectable markers into hematopoietic cells can be used to select genetically modified cells in vivo or to attenuate the toxic effects of chemotherapeutic agents. Fantz *et al.* ((1998) *Biochem Biophys Res Comm* 243(1):6-12) have shown that

30

-77-

retroviral gene transfer of thymidylate synthase (TS) confers resistance to TS directed anticancer agents and that co-expression of TS and dihydrofolate reductase (DHFR) confers resistance to TS and DHFR cytotoxic agents. Retroviral vectors encoding *Escherichia coli* TS, human TS, and the Tyr-to-His at residue 33 variant of human TS (Y33HhTS) were constructed and fibroblasts transfected with these vectors conferred comparable resistance to the TS-directed agent fluorodeoxyuridine (FdUrd, approximately 4-fold). Retroviral vectors that encode dual expression of Y33HhTS and the human L22Y DHFR (L22YhDHFR) variants conferred resistance to FdUrd (3- to 5-fold) and trimetrexate (30- to 140-fold). A L22YhDHFR-Y33HhTS chimeric retroviral vector was also constructed and transduced cells were resistant to FdUrd (3-fold), AG337 (3-fold), trimetrexate (100-fold) and methotrexate (5-fold). These results show that recombinant retroviruses can be used to transfer the cDNA that encodes TS and DHFR and dual expression in transduced cells is sufficiently high to confer resistance to TS and DHFR directed anticancer agents. ACes can be similarly constructed.

Human CD34+ cells do not express glutathione S-transferases alpha

The expression of glutathione S-transferases alpha (GST alpha) in human hematopoietic CD34+ cells and bone marrow was studied using RT-PCR and immunoblotting (Czerwinski M, Kiem *et al.* (1997) *Gene Ther* 4(3):268-70). The GSTA1 protein conjugates glutathione to the stem cell selective alkylator busulfan. This reaction is the major pathway of elimination of the compound from the human body. Human hematopoietic CD34+ cells and bone marrow do not express GSTA1 message, which was present at a high level in liver, an organ relatively resistant to busulfan toxicity in comparison to bone marrow. Similarly, baboon CD34+ cells and dog bone marrow do not express GSTA1. Thus, human

-78-

GSTA1 is a chemoprotective selectable marker in human stem cell gene therapy and could be employed in ACes construction.

Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer

5 Pawliuk *et al.* ((1994) *Blood* 84(9):2868-2877) have investigated the use of a cell surface antigen as a dominant selectable marker to facilitate the detection and selection of retrovirally infected target cells. The small coding region of the human cell surface antigen CD24 (approximately 240 bp) was introduced into a myeloproliferative sarcoma virus (MPSV)-based retroviral vector, which was then used to infect day 4 10 5-fluorouracil (5-FU)-treated murine bone marrow cells. Within 48 hours of termination of the infection procedure CD24-expressing cells were selected by fluorescent-activated cell sorting (FACS) with an antibody directed against the CD24 antigen. Functional analysis of these cells 15 showed that they included not only *in vitro* clonogenic progenitors and day 12 colony-forming unit-spleen but also cells capable of competitive long-term hematopoietic repopulation. Double-antibody labeling studies performed on recipients of retrovirally transduced marrow cells showed that some granulocytes, macrophages, erythrocytes, and, to a lesser 20 extent, B and T lymphocytes still expressed the transduced CD24 gene at high levels 4 months later. No gross abnormalities in hematopoiesis were detected in mice repopulated with CD24-expressing cells. These results show that the use of the CD24 cell surface antigen as a retrovirally encoded marker permits rapid, efficient, and nontoxic selection *in vitro* of 25 infected primary cells, facilitates tracking and phenotyping of their progeny, and provides a tool to identify elements that regulate the expression of transduced genes in the most primitive hematopoietic cells. ACes could be similarly constructed.

-79-

DeltahGHR, a biosafe cell surface-labeling molecule for analysis and selection of genetically transduced human cells

A selectable marker for retroviral transduction and selection of human and murine cells is known (see, Garcia-Ortiz *et al.* (2000) *Hum Gene Ther* 11(2):333-46). The molecule expressed on the cell surface of the transduced population is a truncated version of human growth hormone receptor (deltahGHR), capable of ligand (hGH) binding, but devoid of the domains involved in signal triggering. The engineered molecule is stably expressed in the target cells as an inert protein unable to trigger proliferation or to rescue the cells from apoptosis after ligand binding. This new marker, has a wide application spectrum, since hGHR in the human adult is highly expressed only in liver cells, and lower levels have been reported in certain lymphocyte cell populations. The deltagHHR label has high biosafety potential, as it belongs to a well-characterized hormonal system that is nonessential in adults, and there is extensive clinical experience with hGH administration in humans. The differential binding properties of several monoclonal antibodies (MAbs) are used in a cell rescue method in which the antibody used to select deltagHHR-transduced cells is eluted by competition with hGH or, alternatively biotinylated hGH is used to capture tagged cells. In the latter system, the final purified population is recovered free of attached antibodies in hGH (a substance approved for human use)-containing medium. Such a system could be used to identify ACes containing cells.

4. Transgenic models for evaluation of genes and discovery of new traits in plants

Of interest is the use of plants and plant cells containing artificial chromosomes for the evaluation of new genetic combinations and discovery of new traits. Artificial chromosomes, by virtue of the fact that they can contain significant amounts of DNA can also therefore encode

-80-

numerous genes and accordingly a multiplicity of traits. It is contemplated here that artificial chromosomes, when formed from one plant species, can be evaluated in a second plant species. The resultant phenotypic changes observed, for example, can indicate the nature of the genes contained within the DNA contained within the artificial chromosome, and hence permit the identification of novel genetic activities. Artificial chromosomes containing euchromatic DNA or partially containing euchromatic DNA can serve as a valuable source of new traits when transferred to an alien plant cell environment. For example, it is contemplated that artificial chromosomes derived from dicot plant species can be introduced into monocot plant species by transferring a dicot artificial chromosome. The dicot artificial chromosome possessing a region of euchromatic DNA containing expressed genes.

The artificial chromosomes can be designed to allow the artificial chromosome to recombine with the naturally occurring plant DNA in such a fashion that a large region of naturally occurring plant DNA becomes incorporated into the artificial chromosome. This allows the artificial chromosome to contain new genetic activities and hence carry novel traits. For example, an artificial chromosome can be introduced into a wild relative of a crop plant under conditions whereby a portion of the DNA present in the chromosomes of the wild relative is transferred to the artificial chromosome. After isolation of the artificial chromosome, this naturally occurring region of DNA from the wild relative, now located on the artificial chromosome can be introduced into the domesticated crop species and the genes encoded within the transferred DNA expressed and evaluated for utility. New traits and gene systems can be discovered in this fashion. The artificial chromosome can be modified to contain sequences that promote homologous recombination within plant cells, or

-81-

be modified to contain a genetic system that functions as a site-specific recombination system.

Artificial chromosomes modified to recombine with plant DNA offer many advantages for the discovery and evaluation of traits in different plant species. When the artificial chromosome containing DNA from one plant species is introduced into a new plant species, new traits and genes can be introduced. This use of an artificial chromosome allows for the ability to overcome the sexual barrier that prevents transfer of genes from one plant species to another species. Using artificial chromosomes in this fashion allows for many potentially valuable traits to be identified including traits that are typically found in wild species. Other valuable applications for artificial chromosomes include the ability to transfer large regions of DNA from one plant species to another, such as DNA encoding potentially valuable traits such as altered oil, carbohydrate or protein composition, multiple genes encoding enzymes capable of producing valuable plant secondary metabolites, genetic systems encoding valuable agronomic traits such as disease and insect resistance, genes encoding functions that allow association with soil bacterium such as growth promoting bacteria or nitrogen fixing bacteria, or genes encoding traits that confer freezing, drought or other stress tolerances. In this fashion, artificial chromosomes can be used to discover regions of plant DNA that encode valuable traits.

The artificial chromosome can also be designed to allow the transfer and subsequent incorporation of these valuable traits now located on the artificial chromosome into the natural chromosomes of a plant species. In this fashion the artificial chromosomes can be used to transfer large regions of DNA encoding traits normally found in one plant species into another plant species. In this fashion, it is possible to derive a plant cell that no longer needs to carry an artificial chromosome to

-82-

posses the novel trait. Thus, the artificial chromosome would serve as the transfer mechanism to permit the formation of plants with greater degree of genetic diversity.

The design of an artificial chromosome to accomplish the afore-
5 mentioned purposes can include within the artificial chromosome the presence of specific DNA sequences capable of acting as sites for homologous recombination to take place. For example, the DNA sequence of *Arabidopsis* is now known. To construct an artificial chromosome capable of recombining with a specific region of *Arabidopsis*
10 DNA, a sequence of *Arabidopsis* DNA, normally located near a chromosomal location encoding genes of potential interest can be introduced into an artificial chromosome by methods provided herein. It may be desirable to include a second region of DNA within the artificial chromosome that provides a second flanking sequence to the region
15 encoding genes of potential interest, to promote a double recombination event which would ensure transfer of the entire chromosomal region, encoding genes of potential interest, to the artificial chromosome. The modified artificial chromosome, containing the DNA sequences capable of homologous recombination region, can then be introduced into
20 *Arabidopsis* cells and the homologous recombination event selected.

It is convenient to include a marker gene to allow for the selection of a homologous recombination event. The marker gene is preferably inactive unless activated by an appropriate homologous recombination event. For example, US 5,272,071, describes a method where an
25 inactive plant gene is activated by a recombination event such that desired homologous recombination events can be easily scored. Similarly, US 5,501,967 describes a method for the selection of homologous recombination events by activation of a silent selection gene first introduced into the plant DNA, the gene being activated by an appropriate

-83-

homologous recombination event. Both of these methods can be applied to enable a selective process to be included to select for recombination between an artificial chromosome and a plant chromosome. Once the homologous recombination event is detected, the artificial chromosome, once selected, is isolated and introduced into a recipient cell, for example, tobacco, corn, wheat or rice, and the expression of the newly introduced DNA sequences evaluated.

Phenotypic changes in the recipient plant cells containing the artificial chromosome, or in regenerated plants containing the artificial chromosome, allows for the evaluation of the nature of the traits encoded by the *Arabidopsis* DNA, under conditions naturally found in plant cells, including the naturally occurring arrangement of DNA sequences responsible for the developmental control of the traits in the normal chromosomal environment.

Traits such as durable fungal or bacterial disease resistance, new oil and carbohydrate compositions, valuable secondary metabolites such as phytosterols, flavonoids, efficient nitrogen fixation or mineral utilization, resistance to extremes of drought, heat or cold are all found within different populations of plant species and are often governed by multiple genes. The use of single gene transformation technologies does not permit the evaluation of the multiplicity of genes controlling many valuable traits. Thus, incorporation of these genes into artificial chromosomes allows the rapid evaluation of the utility of these genetic combinations in heterologous plant species.

The large scale order and structure of the artificial chromosome provides a number of unique advantages in screening for new utilities or novel phenotypes within heterologous plant species. The size of new DNA that can be carried by an artificial chromosome can be millions of base pairs of DNA, representing potentially numerous genes that may

-84-

have novel utility in a heterologous plant cell. The artificial chromosome is a "natural" environment for gene expression, the problems of variable gene expression and silencing seen for genes transferred by random insertion into a genome should not be observed. Similarly, there is no
5 need to engineer the genes for expression, and the genes inserted would not need to be recombinant genes. Thus, one expects the expression from the transferred genes to be temporal and spatial, as observed in the species from where the genes were initially isolated. A valuable feature for these utilities is the ability to isolate the artificial chromosomes and to
10 further isolate, manipulate and introduce into other cells artificial chromosomes carrying unique genetic compositions.

Thus, the use of artificial chromosomes and homologous recombination in plant cells can be used to isolate and identify many valuable crop traits.

15 In addition to the use of artificial chromosomes for the isolation and testing of large regions of naturally occurring DNA, methods for the use of artificial chromosomes and cloned DNA are also contemplated. Similar to that described above, artificial chromosomes can be used to carry large regions of cloned DNA, including that derived from other plant species.

20 The ability to incorporate novel DNA elements into an artificial chromosome as it is being formed allows for the development of artificial chromosomes specifically engineered as a platform for testing of new genetic combinations, or "genomic" discoveries for model species such as *Arabidopsis*. It is known that specific "recombinase" systems can be
25 used in plant cells to excise or re-arrange genes. These same systems can be used to derive new gene combinations contained on an artificial chromosome.

The artificial chromosomes can be engineered as platforms to accept large regions of cloned DNA, such as that contained in Bacterial

-85-

Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs). It is further contemplated, that as a result of the typical structure of artificial chromosomes containing tandemly repeated DNA blocks, that sequences other than cloned DNA sequence can be introduced by

5 recombination processes. In particular recombination within a predefined region of the tandemly repeated DNA within the artificial chromosome provides a mechanism to "stack" numerous regions of cloned DNA, including large regions of DNA contained within BACs or YACs clones. Thus, multiple combinations of genes can be introduced onto artificial

10 chromosomes and these combinations tested for functionality. In particular, it is contemplated that multiple YACs or BACs can be stacked onto an artificial chromosomes, the BACs or YACs containing multiple genes of complex pathways or multiple genetic pathways. The BACs or YACs are typically selected based on genetic information available within

15 the public domain, for example from the Arabidopsis Information Management System (<http://aims.cps.msu.edu/aims/index.html>) or the information related to the plant DNA sequences available from the Institute for Genomic Research (<http://www.tigr.org>) and other sites known to those skilled in the art. Alternatively, clones can be chosen at

20 random and evaluated for functionality. It is contemplated that combinations providing a desired phenotype can be identified by isolation of the artificial chromosome containing the combination and analyzing the nature of the inserted cloned DNA.

In this regard, it is contemplated that the use of site-specific

25 recombination sequences can have considerable utility in developing artificial chromosomes containing DNA sequences recognized by recombinase enzymes and capable of accepting DNA sequences containing same. The use of site-specific recombination as a means to target an introduced DNA to a specific locus has been demonstrated in

-86-

the art and such methods can be employed. The recombinase systems can also be used to transfer the cloned DNA regions contained within the artificial chromosome to the naturally occurring plant or mammalian chromosomes.

5 As noted herein, many site-specific recombinases are known and can be identified (Kilby *et al.* (1993) *Trends in Genetics* 9:413-418). The three recombinase systems that have been extensively employed include: an activity identified as R encoded by the pSR1 plasmid of *Zygosaccharomyces rouxii*, FLP encoded for the 2 μ m circular plasmid from
10 *Saccharomyces cerevisiae* and *Cre-lox* from the phage P1.

 The integration function of site-specific recombinases is contemplated as a means to assist in the derivation of genetic combinations on artificial chromosomes. In order to accomplish this, it is contemplated that a first step of introducing site-specific recombinase
15 sites into the genome of a plant cell in an essentially random manner is conducted, such that the plant cell has one or more site-specific recombinase recognition sequences on one or more of the plant chromosomes. An artificial chromosome is then introduced into the plant cell, the artificial chromosome engineered to contain a recombinase
20 recognition site (e.g., integration site) capable of being recognized by a site-specific recombinase. Optionally, a gene encoding a recombinase enzyme is also included, preferably under the control of an inducible promoter. Expression of the site-specific recombinase enzyme in the plant cell, either by induction of a inducible recombinase gene, or
25 transient expression of a recombinase sequence, causes a site-specific recombination event to take place, leading to the insertion of a region of the plant chromosomal DNA (containing the recombinase recognition site) into the recombinase recognition site of the artificial chromosome, and forming an artificial chromosome containing plant chromosomal DNA.

-87-

The artificial chromosome can be isolated and introduced into a heterologous host, preferably a plant host, and expression of the newly introduced plant chromosomal DNA can be monitored and evaluated for desirable phenotypic changes. Accordingly, carrying out this

5 recombination with a population of plant cells wherein the chromosomally located recombinase recognition site is randomly scattered throughout the chromosomes of the plant, can lead to the formation of a population of artificial chromosomes, each with a different region of plant chromosomal DNA, and each potentially representing a novel genetic combination.

10 This method requires the precise site-specific insertion of chromosomal DNA into the artificial chromosome. This precision has been demonstrated in the art. For example, Fukushige and Sauer ((1992) Proc. Natl. Acad. Sci. USA, 89:7905-7909) demonstrated that the *Cre-lox* homologous recombination system could be successfully employed to
15 introduce DNA into a predefined locus in a chromosome of mammalian cells. In this demonstration a promoter-less antibiotic resistance gene modified to include a *lox* sequence at the 5' end of the coding region was introduced into CHO cells. Cells were re-transformed by electroporation with a plasmid that contained a promoter with a *lox* sequence and a
20 transiently expressed *Cre* recombinase gene. Under the conditions employed, the expression of the *Cre* enzyme catalyzed the homologous recombination between the *lox* site in the chromosomally located promoter-less antibiotic resistance gene, and the *lox* site in the introduced promoter sequence, leading to the formation of a functional antibiotic
25 resistance gene. The authors demonstrated efficient and correct targeting of the introduced sequence, 54 of 56 lines analyzed corresponded to the predicted single copy insertion of the DNA due to *Cre* catalyzed site-specific homologous recombination between the *lox* sequences.

-88-

Accordingly a *lox* sequence may be first added to a genome of a plant species capable of being transformed and regenerated to a whole plant to serve as a recombinase target DNA sequence for recombination with an artificial chromosome. The *lox* sequence may be optimally
5 modified to further contain a selectable marker which is inactive but can be activated by insertion of the *lox* recombinase recognition sequence into the artificial chromosome.

A promoterless marker gene or selectable marker gene linked to the recombinase recognition sequence, which is first inserted into the
10 chromosomes of a plant cell can be used to engineer a platform chromosome. A promoter is linked to a recombinase recognition site, in an orientation that allows the promoter to control the expression of the marker or selectable marker gene upon recombination within the artificial chromosome. Upon a site-specific recombination event between a
15 recombinase recognition site in a plant chromosome and the recombinase recognition site within the introduced artificial chromosome, a cell is derived with a recombined artificial chromosome, the artificial chromosome containing an active marker or selectable marker activity that permits the identification and or selection of the cell.

20 The artificial chromosomes can be transferred to other plant or animal species and the functionality of the new combinations tested. The ability to conduct such an inter-chromosomal transfer of sequences has been demonstrated in the art. For example, the use of the *Cre-lox* recombinase system to cause a chromosome recombination event
25 between two chromatids of different chromosomes has been shown.

Any number of recombination systems may be employed as described herein, such as, but not limited to, bacterially derived systems such as the att/int system of phage lambda, and the Gin/gix system.

-89-

More than one recombination system may be employed, including, for example, one recombinase system for the introduction of DNA into an artificial chromosome, and a second recombinase system for the subsequent transfer of the newly introduced DNA contained within an
5 artificial chromosome into the naturally occurring chromosome of a second plant species. The choice of the specific recombination system used will be dependent on the nature of the modification contemplated.

By having the ability to isolate an artificial chromosome, in particular, artificial chromosomes containing plant chromosomal DNA
10 introduced via site-specific recombination, and re-introduce the chromosome into other mammalian or plant cells, particularly plant cells, these new combinations can be evaluated in different crop species without the need to first isolate and modify the genes, or carry out multiple transformations or gene transfers to achieve the same
15 combination isolation and testing combinations of the genes in plants. The use of a site-specific recombinase also allows the convenient recovery of the plant chromosomal region into other recombinant DNA vectors and systems, such as mammalian or insect systems, for manipulation and study.

20 Also contemplated herein are *ACes*, cell lines and methods for use in screening a new chromosomal combinations, deletions, truncations with eucaryotic genome that take advantage of the site-specific recombination systems incorporated onto platform *ACes* provided herein. For example, provided herein is a cell line useful for making a library of
25 *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes. Also provided herein is a method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into a cell line comprising a multiplicity of heterologous recombination

-90-

- sites randomly integrated throughout the endogenous chromosomes, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites within the cell's chromosomal DNA; and isolating said
- 5 multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within. Also provided herein is a library of cells useful for genomic screening, said library comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein. The
- 10 library of cells can be from a different species and/or cell type than the chromosomal nucleic acid within the *ACes*. Also provided is a method of making one or more cell lines, comprising
- a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,
 - 15 b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;
 - c) isolating said multiplicity of *ACes*, thereby producing a library of
 - 20 *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;
 - d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;
 - e) selecting different cells having mutually exclusive *ACes* therein
 - 25 and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

These *ACes*, cell lines and methods utilize the site-specific recombination sites on platform *ACes* analogous YAC manipulation related to: the methods of generating terminal deletions in normal and

-91-

artificial chromosomes (e.g., *ACes*; as described in Vollrath et al., 1988, *PNAS, USA*, 85:6027-66031; and Pavan et al., *PNAS, USA*, 87:1300-1304); the methods of generating interstitial deletions in normal and artificial chromosomes (as described in Campbell et al., 1991, *PNAS, USA*, 88:5744-5748); and the methods of detecting homologous recombination between two *ACes* (as described in Cellini et al., 1991, *Nuc. Acid Res.*, 19(5):997-1000).

5. Use of platform *ACes* in Pharmacogenomic/toxicology applications (development of "Reporter *ACes*")

10 In addition to the placement of genes onto *ACes* chromosomes for therapeutic protein production or gene therapy, the platform can be engineered via the *IntR* lambda integrase to carry reporter-linked constructs (reporter genes) that monitor changes in cellular physiology as measured by the particular reporter gene (or a series of different reporter
15 genes) readout. The reporter linked constructs are designed to include a gene that can be detected (by for example fluorescence, drug resistance, immunohistochemistry, or transcript production, and the like) with well-known regulatory sequences that would control the expression of the detectable gene. Exemplary regulatory promoter sequences are well-
20 known in the art:

A) Reporter *ACes* for drug pathway screening

The *ACes* can be engineered to carry reporter-linked constructs that indicate a signal is being transduced through one or a number of pathways. For example, transcriptionally regulated promoters from genes
25 at the end (or any other chosen point) of particular signal transduction pathways could be engineered on the *ACes* to express the appropriate readout (either by fluorescent protein production or drug resistance) when the pathway is activated (or down-regulated as well). In one embodiment, a number of reporters from different can be placed on a

-92-

ACes chromosome. Cells (and/or whole animals) containing such a Reporter *ACes* could be exposed to a variety of drugs or compounds and monitored for the effects of the drugs or compounds upon the selected pathway(s) by the reporter gene(s). Thus, drugs or compounds can be classified or identified by particular pathways they excite or down-regulate. Similarly, transcriptional profiles obtained from genomic array experiments can be biologically validated using the reporter *ACes* provided herein.

B) Reporter *ACes* for toxic compound testing

Environmental or man-made genotoxigants can be tested in cell lines carrying a number of reporter-genes platform *ACes* linked to promoters that are transcriptionally regulated in response to DNA damage, induced apoptosis or necrosis, and cell-cycle perturbations. Furthermore, new drugs and/or compounds could be tested in a similar manner with the genotoxicant *ACes* reporter for their cellular/genetic toxicity by such a screen. Likewise, toxic compound testing could be carried out in whole transgenic animals carrying the *ACes* chromosome that measures genotoxicant exposure ("canary in a coal mine"). Thus, the same or similar type *ACes* could be used for toxicity testing in either a cell-based or whole animal setting. An example would include *ACes* that carry reporter-linked genes controlled by various cytochrome P450 profiled promoters and the like.

C) Reporter *ACes* for individualized pharmacogenomics/drug profiling

A common disease may arise via various mechanisms. In many instances there are multiple treatments available for a given disease. However, the success of a given treatment may depend upon the mechanism by which the disease originated and/or by the genetic background of the patient. In order to establish the most effective

-93-

treatment for a given patient one could utilize the *ACes* reporters provided herein. *ACes* reporters can be used in patient cell samples to determine an individualized drug regimen for the patient. In addition, potential polymorphisms affecting the transcriptional regulation of an individual's particular gene can be assessed by this approach.

D) Reporter *ACes* for classification of similar patient tumors

As with other diseases as described in 5.C) above, cancer cells arise via different mechanisms. Furthermore, as a cancerous cell propagates it may undergo genomic alterations. An *ACes* reporter transferred to cells of different patients having the same disease, i.e. similar cancers, could be used to categorize the particular cancer of each patient, thereby facilitating the identification of the most effective therapeutic regimen. Examples would include the validation of array profiling of certain classes of breast cancers. Subsequently, appropriate drug profiling could be carried out as described above.

E) Reporter *ACes* as a "differentiation" sensor

Using the *ACes* reporter as a "differentiation" sensor in stem cells or other progenitor cells in order to enrich by selection (either FACS based screening, drug selection and/or use of suicide gene) for a particular class of differentiated or undifferentiated cells. For example, in one embodiment, this assay could also be used for compound screening for small molecule modifiers of cell differentiation.

F) Whole animal studies with Reporter *ACes*

Finally, with whole-body fluorescence imaging technology (Yang et al. (2000) PNAS 97:12278) any of the above Reporter *ACes* methods could be used in conjunction with whole-body imaging to monitor reporter genes within whole animals without sacrificing the animal. This would allow temporal and spatial analysis of expression patterns under a given set of conditions. The conditions tested may include for example, normal

-94-

differentiation of a stem cell, response to drug or compound treatment whether targeted to the diseased tissue or presented systemically, response to genotoxics, and the like.

The following examples are included for illustrative purposes only
5 and are not intended to limit the scope of the invention.

EXAMPLE 1

pFK161

Cosmid pFK161 (SEQ ID NO: 118) was obtained from Dr. Gyula Hadlaczký and contains a 9 kb *NotI* insert derived from a murine rDNA
10 repeat (see clone 161 described in PCT Application Publication No. WO97/40183 by Hadlaczký *et al.* for a description of this cosmid). This cosmid, referred to as clone 161 contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 26. It was produced by inserting fragments of the megachromosome (see, U.S. Patent No.
15 6,077,697 and International PCT application No. WO 97/40183). For example, H1D3, which was deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, is a mouse-hamster hybrid cell line carrying this megachromosome into plasmid pWE15 (Stratagene, La Jolla, California; SEQ ID No. 31) as
20 follows. Half of a 100 μ l low melting point agarose block (mega-plug) containing isolated SATACs was digested with *NotI* overnight at 37°C. Plasmid pWE15 was similarly digested with *NotI* overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 DNA ligase. Ligation was conducted at 16°C overnight. Bacterial
25 DH5 α cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and analyzed by Southern blot hybridization for the presence of DNA that

-95-

hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected.

Liquid cultures of all 189 transformants were used to generate
5 cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones
10 (113, 157 and 161) contained the same insert.

For sequence analysis the insert of cosmid clone no. 161 was subcloned as follows. To obtain the end fragments of the insert of clone no. 161, the clone was digested with *NotI* and *Bam*HI and ligated with *NotI*/*Bam*HI-digested pBluescript KS (Stratagene, La Jolla, California).
15 Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with *Bam*HI-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

20 The insert corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 18. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 19-25. Specifically, the individual
25 subclones corresponded to the following positions in GENBANK accession no. X82564 (SEQ ID NO:18) and in SEQ ID NOS. 19-25:

-96-

5

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<i>NotI, BamHI</i>	19
161m5	7756	8494	<i>BamHI</i>	20
161m7	8495	10231	<i>BamHI</i>	21 (shows only sequence corresponding to nt. 8495-8950), 22 (shows only sequence corresponding to nt. 9851-10231)
161m12	10232	15000	<i>BamHI</i>	23 (shows only sequence corresponding to nt. 10232-10600), 24 (shows only sequence corresponding to nt. 14267-15000)
161k2	15001	15676	<i>NotI, BamHI</i>	25

10 The sequence set forth in SEQ ID NOs. 19-25 diverges in some positions from the sequence presented in positions 7551-15670 of GENBANK accession no. X82564. Such divergence may be attributable to random mutations between repeat units of rDNA.

15 For use herein, the rDNA insert from the clone was prepared by digesting the cosmid with *NotI* and *BglII* and was purified as described above. Growth and maintenance of bacterial stocks and purification of plasmids were performed using standard well known methods (see, *e.g.*, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press), and plasmids were purified from bacterial cultures using Midi - and Maxi-preps Kits (Qiagen, 20 Mississauga, Ontario).

pDsRed1N1

25 This vector is available from Clontech (see SEQ ID No. 29) and encodes the red fluorescent protein (DsRed; Genbank accession no. AF272711; SEQ ID Nos. 39 and 40). DsRed, which has a vivid red fluorescence, was isolated from the IndoPacific sea anemone relative *Discosoma* species. The plasmid pDsRed1N1 (Clontech; SEQ ID No. 29) constitutively expresses a human codon-optimized variant of the

-97-

fluorescent protein under control of the CMV promoter. Unmodified, this vector expresses high levels of DsRed1 and includes sites for creating N-terminal fusions by cloning proteins of interest into the multiple cloning site (MCS). It is Kan and Neo resistant for selection in bacterial or

5 eukaryotic cells.

Plasmid pMG

Plasmid pMG (InvivoGen, San Diego, California; see SEQ. ID. NO. 27 for the nucleotide sequence of pMG) contains the hygromycin phosphotransferase gene under the control of the immediate-early human
10 cytomegalovirus (hCMV) enhancer/promoter with intron A. Vector pMG also contains two transcriptional units allowing for the coexpression of two heterologous genes from a single vector sequence.

The first transcriptional unit of pMG contains a multiple cloning site for insertion of a gene of interest, the hygromycin phosphotransferase
15 gene (*hph*) and the immediate-early human cytomegalovirus (hCMV) enhancer/promoter with intron A (see, *e.g.*, Chapman *et al.* (1991) *Nuc. Acids Res.* 19:3979-3986) located upstream of *hph* and the multiple cloning site, which drives the expression of *hph* and any gene of interest inserted into the multiple cloning site as a polycistronic mRNA. The first
20 transcriptional unit also contains a modified EMCV internal ribosomal entry site (IRES) upstream of the *hph* gene but downstream of the hCMV promoter and MCS for ribosomal entry in translation of the *hph* gene (see SEQ ID NO. 27, nucleotides 2736-3308). The IRES is modified by insertion of the constitutive *E. coli* promoter (EM7) within an intron (IM7)
25 into the end of the IRES. In mammalian cells, the *E. coli* promoter is treated as an intron and is spliced out of the transcript. A polyadenylation signal from the bovine growth hormone (bGh) gene (see, *e.g.*, Goodwin and Rottman (1992) *J. Biol. Chem.* 267:16330-16334) and a pause site derived from the 3' flanking region of the human $\alpha 2$

-98-

globin gene (see, *e.g.*, Enriquez-Harris *et al.* (1991) *EMBO J.* 10:1833-1842) are located at the end of the first transcription unit. Efficient polyadenylation is facilitated by inserting the flanking sequence of the bGh gene 3' to the standard AAUAAA hexanucleotide sequence.

5 The second transcriptional unit of pMG contains another multiple cloning site for insertion of a gene of interest and an EF-1 α /HTLV hybrid promoter located upstream of this multiple cloning site, which drives the expression of any gene of interest inserted into the multiple cloning site. The hybrid promoter is a modified human elongation factor-1 alpha (EF-1
10 alpha) gene promoter (see, *e.g.*, Kim *et al.* (1990) *Gene* 91:217-223) that includes the R segment and part of the U5 sequence (R-U5') of the human T-cell leukemia virus (HTLV) type I long terminal repeat (see, *e.g.*, Takebe *et al.* (1988) *Mol. Cell. Biol* 8:466-472). The Simian Virus 40 (SV40) late polyadenylation signal (see Carswell and Alwine (1989) *Mol.*
15 *Cell. Biol.* 9:4248-4258) is located downstream of the multiple cloning site. Vector pMG contains a synthetic polyadenylation site for the first and second transcriptional units at the end of the transcriptional unit based on the rabbit β -globin gene and containing the AATAAA hexanucleotide sequence and a GT/T-rich sequence with 22-23
20 nucleotides between them (see, *e.g.*, Levitt *et al.* (1989) *Genes Dev.* 3:1019-1025). A pause site derived from the C2 complement gene (see, Moreira *et al.* (1995) *EMBO J.* 14:3809-3819) is also located at the 3' end of the second transcriptional unit.

Vector pMG also contains an ori sequence (ori pMB1) located
25 between the SV40 polyadenylation signal and the synthetic polyadenylation site.

EXAMPLE 2

A. Construction of targeting vector and transfection into LMtk- cells for the generation of platform chromosomes

-99-

A targeting vector derived from the vector pWE15 (GeneBank Accession # X65279) was modified by replacing the *SaI* (Klenow filled)/*SmaI* neomycin resistance containing fragment with the *PvuII*/*Bam*HI (Klenow filled) puromycin resistance containing fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc. Palo Alto, CA; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *NotI* fragment from the plasmid pFK161 (SEQ ID NO: 118) containing a portion of the mouse rDNA region was cloned into the *NotI* site of pWEPuro resulting in plasmid pWEPuro9K (Figure 2). The vector pWEPuro9K was digested with *SpeI* to linearize and transfected into LMtk- mouse cells. Puromycin resistant colonies were isolated and subsequently tested for artificial chromosome formation via fluorescent *in situ* hybridization (FISH) (using mouse major and minor DNA repeat sequences, the puromycin gene and telomeres sequences as probes), and fluorescent activated cell sorting (FACS). From this sort, a subclone was isolated containing an artificial chromosome, designated 5B11.12, which carries 4-8 copies of the puromycin resistance gene contained on the pWEPuro9K vector. FISH analysis of the 5B11.12 subclone demonstrated the presence of telomeres and mouse minor on the *ACes*. DOT PCR has been done on the 5B11.12 *ACes* revealing the absence of uncharacterized euchromatic regions on the *ACes*. A recombination site, such as an *att* or *loxP* engineering site or a plurality thereof, was introduced onto this *ACes* thereby providing a platform for site-specific introduction of heterologous nucleic acid.

B. Targeting a single sequence specific recombination site onto platform chromosomes

After the generation of the 5B11.12 platform, a single sequence-specific recombination site is placed onto the platform chromosome via homologous recombination. For this, DNA sequences containing the site-

-100-

specific recombination sequence can be flanked with DNA sequences of homology to the platform chromosome. For example, using the platform chromosome made from the pWEPuro9K vector, mouse rDNA sequences or mouse major satellite DNA can be used as homologous sequences to

5 target onto the platform chromosome. A vector is designed to have these homologous sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the platform chromosome, the DNA is transfected into cells harboring the platform chromosome (Figure 3). Examples of site-specific

10 cassettes that are targeted to the platform chromosome using either mouse rDNA or mouse major repeat DNA include the SV40-attP-hygro cassette and a red fluorescent protein (RFP) gene flanked by loxP sites (Cre/lox, see, *e.g.*, U.S. Patent No. 4,959,317 and description herein). After transfection and integration of the site-specific cassette,

15 homologous recombination events onto the platform chromosome are subcloned and identified by FACS (*e.g.* screen and single cell subclone via expression of resistance or fluorescent marker) and PCR analysis.

For example, a vector can be constructed containing regions of the mouse rDNA locus flanking a gene cassette containing the SV40 early

20 reporter-bacteriophage lambda attP site-hygromycin selectable marker (see Figure 4 and described below). The use of the bacteriophage lambda attP site for lambda integrase-mediated site-specific recombination is described below. Homologous recombination event of the SV40-attP-hygro cassette onto the platform chromosome was identified using PCR

25 primers that detect the homologous recombination and further confirmed by FISH analysis. After identifying subcloned colonies containing the platform chromosome with a single site-specific recombination site, cells carrying the platform chromosome with a single site-specific

-101-

recombination site can now be engineered with site-specific recombinases (e.g. lambda INT, Cre) for integrating a target gene expression vector.

**C. Targeting a red fluorescent protein (RFP) gene flanked by loxP sites
5 onto 5B11.12 platform**

As another example, while loxP recombination sites could have been introduced onto the *ACes* during *de novo* biosynthesis, it was thought that this might result in multiple segments of the *ACes* containing a high number of loxP sites, potentially leading to instability upon Cre-mediated recombination. A gene targeting approach was therefore
10 devised to introduce a more limited number of loxP recombination sites into a locus of the 5B11-12 *ACes* containing introduced and possibly co-amplified endogenous rDNA sequences. Although there are more than 200 copies of rDNA genes in the haploid mouse genome distributed
15 amongst 5-11 chromosomes (depending on strain), rDNA sequences were chosen as the target on the *ACes* since they represent a less frequent target than that of the satellite repeat sequences. Moreover, having observed much stronger pWEPuro9K hybridization to the 5B11-12 *ACes* than to other LMTK⁻ chromosomes and in light of the observation that the
20 transcribed spacer sequences within the rDNA may be less conserved than the rRNA coding regions, it was contemplated that a targeting vector based on the rDNA gene segment in pWEPuro9K would have a higher probability of targeting to the *ACes* rather than to other LMTK⁻ chromosomes. Accordingly, a targeting vector, pBSFKLoxDsRedLox, was
25 designed and constructed based on the rDNA sequences contained in pWEPuro9K.

The plasmid pBSFKLoxDsRedLox was generated in 4 steps. First, the *NotI* rDNA insert of pWEPuro9K (Figure 2) was inserted into pBS SK- (Stratagene) giving rise to pBSFK. Second, a loxP polylinker cassette was

-102-

generated by PCR amplification of pNEB193 (SEQ ID NO:32; New England Biolabs) using primers complementary to the M13 forward and reverse priming sites at their 3' end and a 34 bp 5' extension comprising a LoxP site. This cassette was reinserted into pNEB193 generating
5 p193LoxMCSLox. Third, the DsRed gene from pDsRed1-N1 (SEQ ID NO:29; Clontech) was then cloned into the polylinker between the loxP sites generating p193LoxDsRedLox. Fourth, a fragment consisting of the DsRed gene flanked by loxP sites was cloned into a unique *NdeI* within the rDNA insert of pBSFK generating pBSFKLoxDsRedLox.

10 A gel purified 11 Kb *PmlI* /*EcoRV* fragment of pBSFKLoxDsRedLox was used for transfection. To detect targeted integration, PCR primers were designed from rDNA sequences within the 5' *NotI*-*PmlI* fragment of pWEPuro9K that is not present on the targeting fragment (5' primer) and sequence within the LoxDsRedLox cassette (3' primer). If the targeting
15 DNA integrated correctly within the rDNA sequences, PCR amplification using these primers would give rise to a 2.3 Kb band. PCR reactions containing 1-4 μ l of genomic DNA were carried out according to the MasterTaq protocol (Eppendorf), using murine rDNA 5' primer (5'-CGGACAATGCGGTTGTGCGT-3'; SEQ ID NO:72) and DsRed 3' primer
20 (5'-GGCCCCGTAATGCAGAAGAA-3'; SEQ ID NO:73) and PCR products were analyzed by agarose gel electrophoresis.

1.5X10⁶ 5B11-12 LMTK⁻ cells were transfected with 2 μ g of the pBSFKLoxDsRedLox targeting DNA described above using Lipofectamine Plus (Invitrogen). For flow sorting, harvested cells were suspended in
25 medium and applied to the Becton Dickinson Vantage SE cell sorter, equipped with 488 nm lasers for excitation and 585/42 bandpass filter for optimum detection of RFP fluorescence. Cells were sorted using dPBS as sheath buffer. Negative control parental 5B11-12 cells and a positive control LMTK⁻ cell line stably transfected with DsRed were used to

-103-

establish the selection gates. The RFP positive gated populations were recovered, diluted in medium supplemented with 1X penicillin-streptomycin (Invitrogen), then plated and cultured as previously described. After 4 rounds of enrichment, the percentage of RFP positive
5 cells reached levels of 50% or higher. DNA from populations was analyzed by PCR for evidence of targeted integration. Ultimately, single cell subclones were established from positive pools and were analyzed by PCR and PCR-positive clones confirmed by FISH as described below. DNA was purified from pools or single cell clones using previously
10 described methods set forth in Lahm et al., Transgenic Res., 1998; 7:131-134, or in some cases using a Wizard Genomic DNA purification kit (Promega). For FISH analysis, a biotinylated DsRed gene probe was generated by PCR using DsRed specific primers and biotin-labeled dUTP (5' RFP primer: 5'-GGTTTAAAGTGCGCTCCTCCAAGAACGTCATC-3',
15 SEQ ID NO:74; and 3' RFP primer: 5'AGATCTAGAGCCGCGCTACAGGAACAGGTGGTGGCGGCC-3'; SEQ ID NO:75). To maximize the signal intensity of the DsRed probe, Tyramide amplification was carried out according to the manufacturers protocols (NEN).

20 The process of testing the feasibility of a more general targeting strategy that would not rely on enrichment *via* drug selection of stably transfected clones can be summarized as follows. A red fluorescent protein gene (RFP; encoded by the DsRed gene) was inserted between the loxP sites of the targeting vector to form pBSFKLoxDsRedLox. After
25 transfection with pBSFKLoxDsRedLox, sequential rounds of high speed flow sorting and expansion of sorted cells in culture could then be used to enrich for stable transformants expressing RFP. In the event of targeted integration, PCR screening with primers that amplify from a spacer region within the segment of the 45s pre-rRNA gene in pWEPuro9K to a specific

-104-

anchor sequence within the DsRed gene in the targeting cassette would give rise to a diagnostic 2.3 Kb band. However, as rDNA clusters are found on several chromosomes, confirmation of targeting to an *ACes* would require fluorescence in situ hybridization (FISH) analysis. Finally, the flanking of the DsRed gene by loxP sites would allow for its removal and subsequent replacement with other genes of interest.

After transfection of the targeting sequence into 5B11-12 cells, enrichment for targeted clones was carried out using a combination of flow cytometry to detect red-fluorescing cells and PCR screening.

10 Ultimately 17 single cell subclones were identified as potential targeted clones by PCR and of these 16 were found by FISH to contain the DsRed integration event into the *ACes*. These subclones are referred to herein as D11-C4, D11-C12, D11-H3, C9-C9, C9-B9, C9-F4, C9-H8, C9-F2, C9-G8, C9-B6, C9-G3, C9-E12, C9-A11, C11-E3, C11-A9 and C11-H4. PCR

15 analysis of genomic DNA isolated from the D11-C4 subclone gave rise to a 2.3 Kb band, indicative of a targeted integration into an rDNA locus. Further analysis of the subclone by FISH analysis with a DsRed gene probe demonstrated integration of the LoxDsRedLox targeting cassette on the *ACes* co-localizing with one of the regions of rDNA staining seen on

20 the 5B11-12 *ACes*, consistent with a targeted integration into an rDNA locus of the *ACes*, while integrations on other chromosomes were not observed. Since transfected cells were maintained as heterogeneous populations through several cycles of sorting and replating it was not possible to estimate the frequency of targeted events. In most

25 mammalian cell lines the frequency of gene targeting via homologous recombination is roughly 10^{-5} - 10^{-7} treated cells. Despite the low frequency of these events in mammalian cells, it is clear that an RFP expression based screening paradigm, coupled with PCR analysis, can effectively detect and enrich for such infrequent events in a large

-105-

population. In instances where drug selection is not possible or not desirable, such a system may provide a useful alternative. It was also verified that the modified *ACes* in subclone D11-C4 could be purified by flow cytometry. The results indicate that the flow karyogram of the D11-C4 subclone was unaltered from that of the 5B11-12 cell line. Thus, the D11-C4 *ACes* can be purified in high yield from native chromosomes of the host cell line.

D. Reduction of LoxP on *ACes* to a single site.

The strong hybridization signal detected by FISH on the *ACes* using the DsRed gene probe suggests that several copies of the targeting cassette may be present on the *ACes* in the D11-C4 line. This also suggests that multiple rDNA genes have been correctly targeted.

Accordingly, in certain embodiments where necessary, the number of loxP sites on the *ACes* can be reduced to a single site by *in situ* treatment with *Cre* recombinase, provided that the sites are co-linear. Such a process is described for multiple loxP-flanked integrations on a native mouse chromosome (Garrick et al., Nature Genet., 1998, Jan;18(1):56-59). Reduction to a single loxP site on the D11-C4 *ACes* would result in the loss of the DsRed gene, forming the basis of a useful screen for this event.

For this purpose, a Cre expression plasmid pCX-Cre/GFP III has been generated by first deleting the EcoRI fragment of pCX-eGFP (SEQ ID NO:71) containing the eGFP coding sequence and replacing it with that of a PCR amplified Cre recombinase coding sequence (SEQ ID NO:58), generating pCX-Cre. Next, the *Asel*/*Sspl* fragment of pD2eGFP-N1 (containing the CMV promoter driving the D2EGFP gene with SV40 polyA signal; Clontech; SEQ ID NO:87) was inserted into the filled *HindIII* site of pCX-Cre, generating pCX-Cre\GFP III. Control plasmid pCX-CreRev\GFP

-106-

III was generated in similar fashion except that the Cre recombinase coding sequence was inserted in the antisense orientation. LMTK⁻ cell line D11-C4 (containing first generation platform ACes with multiple loxP-DsRED sites) and 5B11-12 cell line (containing ACes with no loxP-DsRED sites) are maintained in culture as described above. D11C4 cells are transfected with 2 μ g of plasmid pCX-Cre\GFP III or 2 μ g pCX-CreRev\GFP III using Lipofectamine (Invitrogen) as previously described.

Forty-eight to seventy-two hours after transfection, transfected D11-C4 cells are harvested and GFP positive cells are sorted by cell cytometry using a FACSta Vantage cell sorter (Beckton-Dickinson) as follows: All D11-C4 cells transfected with pCX-Cre\GFP III or control plasmid pCX-CreRev\GFP III that exhibit GFP fluorescent higher than the gate level established by untransfected cells are collected and placed in culture a further 7-14 days. After 7-14 days the initial D11-C4 cells are harvested and analyzed by cell cytometry as follows: Untransfected D11-C4 cells are used to establish the gate that defines the RFP positive population, while 5B11-12 cells are used to set the RFP negative gate. The GFP positive population of D11-C4 transfected with pCX-Cre\GFP III should show decreased red fluorescence compared to pCX-CreRev\GFP III transfected or untransfected control D11-C4 cells. The cells exhibiting greatly decreased or no RFP expression are collected and single cell clones subsequently established. These clones will be expanded and analyzed by fluorescence *in-situ* hybridization and Southern blotting to confirm the removal of loxP-DsRed gene copies.

25

EXAMPLE 3

Construction of targeting vector and transfection into LMTK⁻ cells for the generation of platform chromosomes containing multiple site-specific recombination sites

-107-

An example of a selectable marker system for the creation of a chromosome-based platform is shown in Figure 4. This system includes a vector containing the SV40 early promoter immediately followed by (1) a 282 base pair (bp) sequence containing the bacteriophage lambda attP site and (2) the puromycin resistance marker. Initially a *PvuII/StuI* fragment containing the SV40 early promoter from plasmid pPUR (Clontech Laboratories, Inc., Palo Alto, CA; Seq ID No. 30) was subcloned into the *EcoRI/CRI* site of pNEB193 (a PUC19 derivative obtained from New England Biolabs, Beverly, MA; SEQ ID No. 32) generating the plasmid pSV40193. The only differences between pUC19 and pNEB193 are in the polylinker region. A unique *Ascl* site (GGCGCGCC) is located between the *BamHI* site and the *SmaI* site, a unique *PacI* site (TTAATTAA) is located between the *BamHI* site and the *XbaI* site and a unique *PmeI* site (GTTTAAAC) is located between the *PstI* site and the *SaI* site.

The attP site was PCR amplified from lambda genome (GenBank Accession # NC 001416) using the following primers:
attPUP: CCTTGCGCTAATGCTCTGTTACAGG SEQ ID No. 1
attPDWN: CAGAGGCAGGGAGTGGGACAAAATTG SEQ ID No. 2

After amplification and purification of the resulting fragment, the attP site was cloned into the *SmaI* site of pSV40193 and the orientation of the attP site was determined by DNA sequence analysis (plasmid pSV40193attP). The gene encoding puromycin resistance (Puro) was isolated by digesting the plasmid pPUR (Clontech Laboratories, Inc. Palo Alto, CA) with *AgeI/BamHI* followed by filling in the overhangs with Klenow and subsequently cloned into the *Ascl* site downstream of the attP site of pSV40193attP generating the plasmid pSV40193attPsensePUR (Figure 4; SEQ ID NO:113)).

-108-

The plasmid pSV40193attPsensePUR was digested with *ScaI* and co-transfected with the plasmid pFK161 (SEQ ID NO: 118) into mouse LMTk- cells and platform artificial chromosomes were identified and isolated as described above. The process for generating this exemplary platform ACes containing multiple site-specific recombination sites is summarized in Figure 5. One platform ACes resulting from this experiment is designated B19-18. This platform ACes chromosome may subsequently be engineered to contain target gene expression nucleic acids using the lambda integrase mediated site-specific recombination system as described herein in Example 7 and 8.

EXAMPLE 4

Lambda integrase mediated site-specific recombination of a RFP expressing vector onto artificial chromosomes

In this example, a vector expressing the red fluorescent protein (RFP) was produced and recombined into the attP site residing on an artificial chromosome within LMTK- cells. This recombination is depicted in Figure 7.

A. Construction of expression vectors containing wildtype and mutant lambda integrase

Mutations at the glutamic acid at position 174 in the lambda integrase protein relaxes the requirement for the accessory protein IHF during recombination and DNA supercoiling *in vitro* (see, Miller *et al.* (1980) *Cell* 20:721-729; Lange-Gustafson *et al.* (1984) *J. Biol. Chem.* 259:12724-12732). Mutations at this site promote attP, attB intramolecular recombination in mammalian cells (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181).

To construct nucleic acid encoding the mutant, lambda integrase was PCR amplified from bacteriophage lambda DNA (cl857 *ind Sam 7*; New England Biolabs) using the following primers:

Lamint1 (SEQ ID No. 3)

-109-

TTCGAATTCATGGGAAGAAGGCGAAGTCATGAGCG)

Lamint2 (SEQ ID No. 4)

(TTCGAATTCTTATTTGATTTCAATTTTGTCCCAC).

- The resulting PCR product was digested with *EcoR* I and cloned into the
- 5 *EcoR* I site of pUC19. Lambda integrase was mutated at amino acid position 174 using QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the following oligos (generating a glutamic acid to arginine change at position 174):

LambdaINTE174R

- 10 (SEQ ID No. 6)

(CGCGCAGCAAAATCTAGAGTAAGGAGATCAAGACTTACGGCTGACG),

LamintR174rev (SEQ ID No. 7)

(CGTCAGCCGTAAGTCTTGATCTCCTTACTCTAGATTTTGCTGCGCG).

The resulting site directed mutant was confirmed by sequence analysis.

- 15 The wildtype and mutant lambda genes were cloned into the *EcoR* I site of pCX creating pCX-LamInt (SEQ ID NO: 127) and pCXLamIntR (Figure 8; SEQ ID NO: 112).

- The plasmid pCX (SEQ ID No. 70) was derived from plasmid pCXeGFP (SEQ ID No. 71). Excision of the *EcoRI* fragment containing the
- 20 eGFP marker generated pCX. To generate plasmid pCXLamIntR (SEQ ID NO: 112) an *EcoRI* fragment containing the lambda integrase E174R (SEQ ID No. 37) mutation was cloned into the *EcoRI* site of pCX, and to generate plasmid pCX-LamInt, an *EcoRI* fragment containing the wild-type lambda integrase was cloned into the *EcoRI* site of pCX.

- 25 **B. Construction of integration vector containing attB and DsRed**

The plasmid pDsRedN1 (Clontech Laboratories, Palo Alto, CA; SEQ ID No. 29) was digested with *Hpa* I and ligated to the following annealed oligos:

attB1 (SEQ ID No. 8)

-110-

(TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA)

attB2 (SEQ ID No. 9)

(TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA)

The resulting vector (pDsRedN1-attB) was confirmed by PCR and

5 sequence analysis.

C. Transfection into LMtk- cells

LM(tk-) cells containing the Prototype A ACes (L1-18; Chromos
Molecular Systems Inc., Burnaby, BC Canada) were co-transfected with
pDsRedN1 or pDsRedN1-attB and either pCXLamInt (SEQ ID NO: 127) or
10 pCXLamIntR (SEQ ID NO: 112) using Lipofectamine Plus Reagent
(LifeTechnologies, Gaithersburg, MD). The transfected cells were grown
in DMEM (LifeTechnologies, Gaithersburg, MD) with 10% FBS (CanSera)
and G418 (CalBiochem) at a concentration of 1 mg/ml.

D. Enrichment by cell sorting

15 The transfected cells were sorted using a FACs Vantage SE cell
sorter (Becton Dickenson) to enrich for cells expressing DsRed. The cells
were excited with a 488 nm Argon laser at 200 watts and cells
fluorescing in the 585/42 detection channel were collected. The sorted
cells were returned to growth medium for recovery and expansion. After
20 three successive enrichments for cells expressing DsRed, single cell
sorting into 96 well plates was performed using the same parameters.
Duplicate plates of the single cell clones were made for PCR analysis.

E. PCR analysis of single cell clones

25 Pools of cells from each row and column of the 96 well plate were
used for DNA isolation. DNA was prepared using a Wizard Genomic DNA
purification kit (Promega Inc, Madison, WI). Nested PCR analysis on the
DNA pools was performed to confirm the site-specific recombination
event using the following primer sets:

-111-

attPdown2 (SEQ ID No. 10)
(TCTTCTCGGGCATAAGTCGGACACC)

CMVen (SEQ ID No. 11)
(CTCACGGGGATTCCAAGTCTCCAC)

5 followed by:

attPdown (SEQ ID No. 12)
(CAGAGGCAGGGAGTGGGACAAAATTG)

CMVen2 (SEQ ID No. 13)
(CAACTCCGCCCCATTGACGCAAATG).

10 The resulting PCR reactions were analyzed by gel electrophoresis and the potential individual clones containing the site-specific recombination event were identified by combining the PCR results of all of the pooled rows and columns for each 96 well plate. The individual clones were then further analyzed by PCR using the following primers that flank the
15 recombination junction. L1for and F1rev flank the attR junction whereas REDfor and L2rev flank the attL junction (see Figure 7):

L1for (SEQ ID No. 14)
AGTATCGCCGAACGATTAGCTCTTCA

20 F1rev (SEQ ID No. 15)
GCCGATTTCGGCCTATTGGTTAAA

REDfor (SEQ ID No. 16)
CCGCCGACATCCCCGACTACAAGAA

L2rev (SEQ ID No. 17)
TTCCTTCGAAGGGGATCCGCCTACC.

25 **F. Sequence analysis of recombination junctions**

PCR products spanning the recombination junction were Topo-cloned into pcDNA3.1D/V5His (Invitrogen Inc., San Diego, CA) and then sequenced by cycle-sequencing. The clones were confirmed to have the correct *attR* and *attL* junctions by cycle sequencing.

30 **G. Fluorescent In Situ Hybridization (FISH)**

The cell lines containing the correct recombination junction sequence were further analyzed by fluorescent *in situ* hybridization (FISH)

-112-

by probing with the DsRed coding region labeled with biotin and visualizing with the Tyramide Signal Amplification system (TSA; NEN Life Science Products). The results indicate that the RFP sequence is present on the ACes.

5 **H. Southern analysis**

Genomic DNA was harvested from the cell lines containing an ACes with the correct recombinant event and digested with *EcoR* I. The digested DNAs were separated on a 0.7% agarose gel, transferred and fixed to a nylon membrane and probed with RFP coding sequences. The
10 result showed that there is an integrated copy of RFP coding sequence in each clone.

EXAMPLE 5

Delivery of a second gene encoding GFP onto the RFP platform ACes

15 **A. Construction of integration vector containing attB and GFP (pD2eGFPIresPuroattB).**

The plasmid pIRESpuro2 (Clontech, Palo Alto, CA; SEQ ID NO: 88) was digested with *EcoRI* and *NotI* then ligated to the D2eGFP *EcoRI-NotI* fragment from pD2eGFP-N1 (Clontech, Palo Alto, CA) to create pD2eGFPIresPuro2. Subsequently, oligos encoding the attB site were
20 annealed and ligated into the *NruI* site of pD2eGFPIresPuro2 to create pD2eGFPIresPuroattB. The orientation of attB in the *NruI* site was determined by PCR.

B. Transfection of LMtk- cells

The LMtk- cells containing the RFP platform ACes produced in
25 Example 4, which has multiple attP sites, were co-transfected with pCXLamIntR and pD2eGFPIresPuroattB using LipofectAMINE PLUS reagent. Five μ g of each vector was placed into a tube containing 750 μ l of DMEM (Dulbecco's modified Eagles Medium). Twenty μ l of the Plus reagent was added to the DNA and incubated at room temperature for 15

-113-

minutes. A mixture of 30 μ l of lipofectamine and 750 μ l DMEM was added to the DNA mixture and incubated an additional 15 minutes at room temperature. The DNA mixture was then added dropwise to approximately 3 million cells attached to a 10cm dish in 5 mls of DMEM.

- 5 The cells were incubated 4 hours (37°C, 5% CO₂) with the DNA-lipid mixture, after which DMEM with 20% fetal bovine serum was added to the dishes to bring the culture medium to 10% fetal bovine serum. The dishes were incubated at 37°C with 5% CO₂.

Plasmid pD2eGFPIresPuroattB has a puromycin gene

- 10 transcriptionally linked to the GFP gene *via* an IRES element. Two days after the transfection the cells were placed in medium containing puromycin at 4 μ g/ml to select for cells containing the pD2eGFPIresPuroattB plasmid integrated into the genome. Twenty-three clones were isolated after 17 days of selection with puromycin. These
- 15 clones were expanded and then analyzed for the presence of the GFP gene on the ACes by 2-color (RFP/biotin & GFP/digoxigenin) TSA-FISH (NEN) according to the manufacturers protocol. Sixteen of the 23 clones produced a positive FISH signal on the ACes with a GFP probe.

EXAMPLE 6

20 Delivery Of ACes Into human Mesenchymal Stem Cells (hMSC)

A. Transfection

- Transfection conditions for the most efficient delivery of the ACes into hMSCs (Cambrex BioWhittaker Product Code PT-2501, lot# F0658, East Rutherford, New Jersey) were assayed using LipofectAMINE PLUS
- 25 and Superfect. One million prototype B ACes, which is a murine derived 60Mb ACes having primarily murine pericentric heterochromatin, and carrying a "payload" containing a hygromycin B selectable marker gene and a *lacZ* reporter gene (see , Telenius et al., 1999, Chrom. Res., 7:3-7; and Kereso et al., 1996, Chrom. Res., 4:226-239; each of which is

-114-

incorporated herein by reference in its entirety), were combined with 1-12 μ l of the transfection agent. In the case of LipofectAMINE PLUS, the PLUS reagent was combined with the ACes for 15 minutes followed by LipofectAMINE for a further 15 minutes. Superfect was complexed for

5 10 minutes at a ratio of 2 μ l Superfect per 1 million ACes. The ACes/transfection agent complex was then applied to 0.5 million recipient cells and the transfection was allowed to proceed according to the manufacturer's protocol. Percent transfected cells was determined on a FACS Vantage flow cytometer with argon laser tuned to 488 nm at

10 200mW and FITC fluorescence collected through a standard FITC 530/30 nm band pass filter. After 24 hours, IdUrd labeled ACes were delivered to human MSCs in the range of 30-50%, varying with transfection agent and dose. ACes delivery curves were generated from data collected in experiments that varied the dose of the transfection reagents. Dose

15 response curves of Superfect and LipofectAMINE PLUS, showing delivery of ACes into recipient hMSCs cells, were prepared, measured by transfer of IdUrd labeled ACes and detected by flow cytometry. Superfect shows maximum delivery in the range of 30-50% at doses greater than 2 μ l per million ACes. LipofectAMINE PLUS has a 42-48% delivery peak around

20 5-8 μ l per million ACes. These dose curves were then correlated with toxicity data to determine the transfection conditions that will allow for highest potential transfection efficiency. Toxicity was determined by a modified plating efficiency assay (de Jong et al., 2001, Chrom. Research, 9:475-485). The population's normalized plating efficiency (at maximum

25 % delivery doses) was in the range of 0.2 - 0.4 for Superfect and 0.5 - 0.6 with LipofectAMINE PLUS.

Due to the transfected population consisting of mixed cell types, flow cytometry allowed for the assessment of ACes delivery into each sub-population and the purification of the target population. Flow profiles

-115-

showing forward scatter (cell size) and side scatter (internal cell granularity) revealed three distinct hMSC populations that were gated into three regions: R3 (small cell region), R4 (medium cell region), R5 (large cell region). Transfection conditions were further optimized by re-analyzing delivery curves and assessing the differences in delivery to each sub-population. Dose response curves of Superfect and LipofectAMINE were prepared showing % delivery to each sub-population represented by the gating on basis of cell size and granularity properties of the mixed population. Three distinct hMSC populations were gated and % delivery dose curves generated. Using Superfect and LipofectAMINE PLUS the overall % delivery increased with cell size (80-90% delivery in large cells). LipofectAMINE PLUS at high doses (8-12 μ l per 1 million ACes) shows an increase in the overall proportion of chromosome transfer to the small population (10-20%). This suggests an advantage to using this transfection agent if the small-undifferentiated cell population is the desired target host cell.

B. Expression from Genes on ACes IN hMSCs

Following the delivery screening process conducted in section (A) above, the most promising results were subjected to further analyses to monitor expression and verify the presence of structurally intact ACes. The transfection conditions employed for these experiments were exactly the same as those that had been used during the screening process. Short-term expression was monitored by transfecting hMSCs with ACes containing a RFP gene (red fluorescent protein) set forth in Example 2C as "D11C4". The unselected population was harvested at 72-96 hours post transfection and % positive fluorescent cells measured by flow cytometry. RFP expression was in the range of 1-20%.

Long term-gene expression was assayed by selecting for hygromycin B resistant cells over a period of 7-10 days. Cytogenetic

-116-

analysis was done to detect presence of intact ACes by Fluorescent *In Situ* hybridization (FISH), where metaphase chromosomes were hybridized to a mouse major satellite-DNA probe (targeting murine pericentric heterochromatin) and a lambda probe (hybridizing to the *lacZ* gene). The human mesenchymal transfected culture could not undergo standard sub-cloning as diffuse colonies form with limited doublings available for expansion. Cytogenetic analysis was performed on the entire population, sampling over a period of 3-10 days post-transfection. The hygromycin resistant population was then blocked in mitosis with colchicine and analyzed for presence of intact ACes by FISH. Preliminary FISH results show approximately 2-8% of the hMSC-transfected population had an intact ACes. This compared to rat skeletal muscle myoblast clones, which were in the range of 60-95%. To increase the % of intact ACes in the hMSC-transfected population an enrichment step can be utilized as described in Example 2C.

C. Differentiation of The hMSCs

In initial experiments where transfected hMSCs cells have been induced to differentiate into adipose or osteocytes, the results indicate that the transfected cells appear to be differentiating at a rate comparable to the untransfected controls and the cultures are lineage specific as tested by microscopic examination, FISH, Oil Red O staining (adipocyte assay), and calcium secretion (osteocyte assay).

Accordingly, these results indicate that the artificial chromosomes (ACes) provided herein can be successfully transferred into hMSC target cells. Targeting MSCs (such as hMSCs) permits gene transfer into cells in an undifferentiated state where the cells are easier to expand and purify. The genetically modified cells can then be differentiated *in vitro* or injected into a site *in vivo* where the microenvironment will induce transformation into specific cell lineages.

-117-

EXAMPLE 7**Delivery of a Promoterless Marker Gene to a Platform ACes**

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

5 **A. Construction of Targeting Vectors.**

The base vector p18attBZeo (3166bp; SEQ ID NO: 114) was constructed by ligating the 1067bp *HindIII-SspI* fragment containing attBZeo, obtained from pLITattBZeo (SEQ ID NO:91), into pUC18 (SEQ ID NO: 122) digested with *HindIII* and *SspI*.

10 1. p18attBZEO-eGFP (6119bp; SEQ ID NO: 126) was constructed by inserting the 2977bp *SpeI-HindIII* fragment from pCXeGFP (SEQ ID NO:71; Okabe, *et al.* (1997) *FEBS Lett* 407:313-319) containing the eGFP gene into p18attBZeo (SEQ ID NO: 114) digested with *HindIII* and *XbaI*.

15 2. p18attBZEO-5'6XHS4eGFP (Figure 10; 7631bp; SEQ ID NO: 116) was constructed by ligating the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal into the *HindIII* site of p18attBZeo (SEQ ID NO: 20 114).

3. p18attBZEO-3'6XHS4eGFP (Figure 11; 7600bp; SEQ ID NO: 115) was created by removing the 5'6XHS4 element from p18attBZeo-(6XHS4)2eGFP (SEQ ID NO: 110). p18attBZeo-(6XHS4)2eGFP was digested with *EcoRV* and *SpeI*, treated with Klenow and religated to form 25 p18attBZeo3'6XHS4eGFP (SEQ ID NO: 115).

4. p18attBZEO-(6XHS4)2eGFP (Figure 12; 9080bp; SEQ ID NO: 110) was created in two steps. First, the *EcoRI-SpeI* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains 6 copies of the HS4 core element was ligated into p18attBZeo (SEQ ID NO: 114)

-118-

digested with *EcoRI* and *XbaI* to create p18attBZeo6XHS4 (4615bp; SEQ ID NO: 117). Next, p18attBZeo6XHS4 was digested with *HindIII* and ligated to the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal.

Table 2

Targeting plasmid	No. zeocin resistant clones	No. clones with expected PCR product size	No. clones with correct sequence at recombination junction
p18attBZEO-eGFP	12	12	NT*
p18attBZEO-5'6XHS4eGFP	11	11	NT
p18attBZEO-3'6XHS4eGFP	11	11	NT
p18attBZEO-(6XHS4)2eGFP	9	9	4/4

*NT = not tested

B. Transfection and Selection with Drug.

The mouse cell line containing the 2nd generation platform ACE, B19-38 (constructed as set forth in Example 3), was plated onto four 10cm dishes at approximately 5 million cells per dish. The cells were incubated overnight in DMEM with 10% fetal calf serum at 37°C and 5% CO₂. The following day the cells were transfected with 5μg of each of the 4 vectors listed in Example 7.A. above and 5μg of pCXLamIntR (SEQ ID NO: 112), for a total of 10μg per 10cm dish. Lipofectamine Plus reagent was used to transfect the cells according to the manufacturers protocol. Two days post-transfection zeocin was added to the medium at 500ug/ml. The cells were maintained in selective medium until colonies formed. The colonies were then ring-cloned (see, e.g., McFarland, 2000, Methods Cell Sci, Mar;22(1):63-66).

C. Analysis of Clones (PCR, SEQUENCING).

-119-

Genomic DNA was isolated from each of the candidate clones with the Wizard kit (Promega) and following the manufacturers protocol. The following primer set was used to analyze the genomic DNA isolated from the zeocin resistant clones: 5PacSV40 -

- 5 CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO:76);
Antisense Zeo - TGAACAGGGTCACGTCGTCC (SEQ ID NO:77). PCR amplification with the above primers and genomic DNA from the site-specific integration of any of the 4 zeocin vectors would result in a 673bp PCR product.
- 10 As set forth in Table 2, of the 4 zeocin resistant candidate clones thusfar analyzed by PCR, all 4 exhibit the correct sequence for a site-specific integration event.

EXAMPLE 8

Integration of a PCR product by site-specific recombination.

- 15 In this example a gene is integrated onto the platform *ACes* by site-specific recombination without cloning said gene into a vector.

A. PCR PRIMER DESIGN.

- 20 PCR primers are designed to contain an attB site at the 5' end of one of the primers in the primer set. The remaining primers, which could be one or more than one primer, do not contain an attB site, but are complementary to sequences flanking the gene or genes of interest and any associated regulatory sequences. In first example, 2 primers (one containing an attB site) are used to amplify a selective gene such as puromycin.

- 25 In a second example as shown in Figure 13, the primer set includes primers 1 & 2 that amplify the GFP gene without amplification of an upstream promoter. Primer 1 contains the attB site at the 5' end of the oligo. Primers 3 & 4 are designed to amplify the IRES-blasticidin DNA sequences from the vector pIRESblasticidin. The 5'end of primer 3

-120-

contains sequences complementary to the 5' end of primer 2 such that annealing can occur between 5' ends of the two primers.

B. PCR REACTION AND SUBSEQUENT LIGATION TO CREATE CIRCULAR MOLECULES FROM THE PCR PRODUCT

- 5 In the first example set forth above in Section A, the two PCR primers are combined with a puromycin DNA template such as pPUR (Clontech), a heat stable DNA polymerase and appropriate conditions for DNA amplification. The resulting PCR product (attB-Puromycin) is then then purified and self-ligated to form a circular molecule.
- 10 In the second example set forth above in Section A, amplification of the GFP gene and IRES-blasticidin sequences is accomplished by combining primers 1 & 2 with DNA template pD2eGFP and primers 3 & 4 with template pIRESblasticidin under appropriate conditions to amplify the desired template. After initial amplification of the two products (attB-GFP
- 15 & IRES-blasticidin) in separate reactions, a second round of amplification using both of the PCR products from the first round of amplification together with primers 1 and 4 amplifies the fusion product attB-GFP-IRES-blasticidin (Figure 13). This technique of using complementary sequences in primer design to create a fusion product is employed in *Saccharomyces*
- 20 *cerevisiae* for allele replacement (Erdeniz *et al* (1997) *Gen Res* 7:1174-1183). The amplified product is then purified from the PCR reaction mixture by standard methods and ligated to form a circular molecule.

C. INTRODUCTION OF PCR PRODUCT ONTO THE ACes USING A RECOMBINASE

- 25 The circular PCR product is then be introduced to the platform ACes using the bacteriophage lambda integrase E174R. The introduction can be performed *in vivo* by transfecting the pCXLamIntR (SEQ ID NO: 112) vector encoding the lambda integrase mutant E174R together with the circularized PCR product into a cell line containing the platform ACE.

-121-

D. SELECTION FOR MARKER GENE

The marker gene (in this case either puromycin, blasticidin or GFP) is used to enrich the population for cells containing the proper integration event. A proper integration event in the second example (Figure 14) juxtaposes a promoter residing on the platform ACes 5' to the attB-GFP-IRES-Blasticidin PCR product, allowing for transcription of both GFP and blasticidin. If enrichment is done by drug selection, blasticidin is added to the medium on the transfected cells 24-48 hours post-transfection. Selection is maintained until colonies are formed on the plates. If enrichment is done by cell sorting, cells are sorted 2-4 days post-transfection to enrich for cells expressing the fluorescent marker (GFP in this case).

E. ANALYSIS OF CLONES

Clonal isolates are analyzed by PCR, FISH and sequence analysis to confirm proper integration events.

EXAMPLE 9**Construction of a human platform ACes "ACE 0.1"****A. CONSTRUCTION OF THE TARGETING VECTOR pPACrDNA**

Genome Systems (IncyteGenomics) was supplied with the primers 5'HETS (GGGCCGAAACGATCTCAACCTATT; SEQ ID NO:78), and 3'HETS (CGCAGCGGCCCTCCTACTC; SEQ ID NO:79), which were used to amplify a 538bp PCR product homologous to nt 9680-10218 of the human rDNA sequences (GenBank Accession No. U13369) and used as a probe to screen a human genomic P1AC (P1 Artificial Chromosome) library constructed in the vector pCYPAC2 (Ioannou *et al.* (1994) *Nat. Genet.* 6(1): 84-89). Genome Systems clone #18720 was isolated in this screen and contains three repeats of human rDNA as assessed by restriction analysis. GS clone #18720, was digested with PmeI, a restriction enzyme unique to a single repeat of the human rDNA (45Kbp),

-122-

and then religated to form pPACrDNA (Figure 15). The insert in pPACrDNA was analyzed by restriction digests and sequence analysis of the 5' and 3' termini. The pPACrDNA, rDNA sequences are homologous to Genbank Accession #U13369, containing an insert of about 45 kB comprising a single repeat beginning from the end of one repeat at ~33980 (relative to the Genbank sequence) through the beginning of the next repeat up to approximately 35120 (the repeat offset from that listed in the GenBank file). Thus, the rDNA sequence is just over 1 copy of the repeat extending from 33980 (+/-10bp) to the end of the first repeat (43Kbp) and continuing into the second repeat to bp 35120 (+/-10bp).

B. TRANSFECTION AND ACes FORMATION.

Five hundred thousand MSU1.1 cells (Morgan et al., 1991, Exp. Cell Res., Nov;197(1):125-136; provided by Dr. Justin McCormick at Michigan State University) were plated per 6cm plate (3 plates total) and allowed to grow overnight. The cells were 70-80% confluent the following day. One plate was transfected with 15µg pPACrDNA (linearized with *Pme* I) and 2µg pSV40attPsensePuro (linearized with *Sca* I; see Example 3). The remaining plates were controls and were transfected with either 20µg pBS (Stratagene) or 20µg pSV40attBsensePuro (linearized with *Sca* I). All three plates were transfected using a CaPO₄ protocol.

C. SELECTION OF PUROMYCIN RESISTANT COLONIES

One day post-transfection the cells were "glycerol shocked" by the addition of PBS medium containing 10% glycerol for 30 seconds. Subsequently, the glycerol was removed and replaced with fresh DMEM. Four days post-transfection selective medium was added. Selective medium contains 1ug/ml puromycin. The transfection plates were maintained at 37°C with 5% CO₂ in selective medium for 2 weeks at which point colonies could be seen on the plate transfected with

-123-

pPACrDNA and pSV40attPsensePuro. The colonies were ring-cloned from the plate on day 17 post-selection and expanded in selective medium for analysis. Only two colonies (M2-2d & M2-2b) were able to proliferate in the selective medium after cloning. No colonies were seen on the control plates after 37 days in selective medium.

D. ANALYSIS OF CLONES

FISH analysis was performed on the candidate clones to detect ACes formation. Metaphase spreads from the candidate clones were probed in multiple probe combinations. In one experiment, the probes used were biotin-labeled human aliphoid DNA (pPACrDNA) and digoxigenin-labeled mouse major DNA (pFK161) as a negative control. Candidate M2-2d was single cell subcloned by flow sorting and the candidate subclones were reanalyzed by FISH. Subclone 1B1 of M2-2d was determined to be a platform ACes and is also designated human Platform ACE 0.1.

EXAMPLE 10

Site-specific integration of a marker gene onto a human platform ACE 0.1

The promoterless delivery method was used to deliver a promoterless blasticidin marker gene onto the human platform ACes with excellent results. The human ACes platform with a promoterless blasticidin marker gene resulted in 21 of 38 blasticidin resistant clones displaying a PCR product of the expected size from the population co-transfected with pLIT38attBBSRpolyA10 and pCXLamIntR (Figure 8; SEQ ID NOs. 111 and 112). Whereas, the population transfected with pBlueScript resulted in 0 blasticidin resistant colonies.

A. CONSTRUCTION OF pLIT38attB-BSRpolyA10 & pLIT38attB-BSRpolyA2.

The vector pLITMUS 38 (New England Biolabs; U.S. Patent No. 5,691,140; SEQ ID NO: 119) was digested with *EcoRV* and ligated to

-124-

two annealed oligomers, which form an attB site (attB1 5'-TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA-3' (SEQ ID NO:8); attB2 5'- TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA-3'; SEQ ID NO:9). This ligation reaction resulted in the vector pLIT38attB (SEQ ID NO: 120).

- 5 The blasticidin resistance gene and SV40 polyA site was PCR amplified with primers: 5BSD (ACCATGAAAACATTTAACATTTCTCAACA; SEQ ID NO:80) and SV40polyA (TTTATTTGTGAAATTTGTGATGCTATTGC; SEQ ID NO:81) using pPAC4 (Frengen, E., *et al.* (2000) Genomics 68 (2), 118-126; GenBank Accession No. U75992) as template. The blasticidin-
- 10 SV40polyA PCR product was then ligated into pLIT38attB at the *Bam*HI site, which was Klenow treated following digestion with *Bam*HI. pLIT38attB-BSDpolyA10 (SEQ ID NO: 111) and pLIT38attB-BSDpolyA2 (SEQ ID NO: 121) are the two resulting orientations of the PCR product ligated into the vector.

15 **B. TRANSFECTION OF MSU1.1 CELLS CONTAINING HUMAN PLATFORM ACE 0.1.**

- MSU1.1 cells containing human platform ACE 0.1 (see Example 9) was expanded and plated to five 10cm dishes with 1.3×10^6 cells per dish. The cells were incubated overnight in DMEM with 10% fetal bovine
- 20 serum, at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each plasmid as set forth in Table 3, for a total of 10µg of DNA per plate of cells transfected (see Table 3) using ExGen 500 *in vitro* transfection reagent (MBI fermentas, cat. no. R0511). The transfection was performed according to the manufacturers protocol.
- 25 Cells were incubated at 37°C with 5% CO₂ in DMEM with 10% fetal bovine serum following the transfection.

-125-

Table 3

Plate #	Plasmid 1	Plasmid 2	No. Bsd ^R Colonies
1	pBS	None	0
2	pCXLamInt	pLIT38attB-BSRpolyA10	16
3	pCXLamIntR	pLIT38attB-BSRpolyA10	40
4	pCXLamInt	pLIT38attB-BSRpolyA2	28
5	pCXLamIntR	pLIT38attB-BSRpolyA2	36

10 C. SELECTION OF BLASTICIDIN RESISTANT CLONES.

Three days following the transfection the cells were split from a 10 cm dish to two 15cm dishes. The cells were maintained in DMEM with 10% fetal bovine serum for 4 days in the 15 cm dishes. Seven days post-transfection blasticidin was introduced into the medium. Stably transfected cells were selected with 1 μ g/ml blasticidin. The number of colonies formed on each plate is listed in Table 3. These colonies were ring-cloned and expanded for PCR analysis. Upon expansion in blasticidin containing medium some clones failed to live and therefore do not have corresponding PCR data.

20 D. PCR ANALYSIS

Thirty-eight of the 40 clones from plate 3 grew after ring-cloning. Genomic DNA was isolated from these clones with the Promega Wizard Genomic cDNA purification kit, digested with *Eco*RI and used as template in a PCR reaction with the following primers: 3BSP – TTAATTTCTGGG TATATTTGAGTGGA (SEQ ID NO:82); 5PacSV40 – CTGTTAATTAAGTGTGGAA TGTGTGTCAGTTAGGGTG (SEQ ID NO:76). The PCR conditions were as follows. 100ng of genomic DNA was

-126-

amplified with 0.5ul Herculase polymerase (Stratagene) in a 50ul reaction that contained 12.5pmole of each primer, 2.5mM of each dNTP, and 1X Herculase buffer (Stratagene). The reactions were placed in a PerkinElmer thermocycler programmed as follows: Initial denaturation at 95°C for 10 minutes; 35 cycles of 94°C for 1 minute, 53°C for 1 minute, 72°C for 1 minute, and 72°C for 1 minute; Final extension for 10 minutes at 72°C; and 4°C hold. If pLIT38attB-BSRpolyA10 integrates onto the human platform ACE 0.1 correctly, PCR amplification with the above primers should yield an 804bp product. Twenty-one of the 38 clones from plate 3 produced a PCR product of the expected 804bp size.

EXAMPLE 11

Delivery of a Vector comprising a Promoterless Marker Gene and a gene encoding a therapeutic product to a Platform ACes

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

A. CONSTRUCTION OF DELIVERY VECTORS

1. Erythropoietin cDNA vector, p18EPOcDNA.

The erythropoietin cDNA was PCR amplified from a human cDNA library (E. Perkins *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96(5): 2204-2209) using the following primers: EPO5XBA - TATCTAGAATGGGGGTGC ACGAATGTCCTGCC (SEQ ID NO: 83); EPO3BSI - TACGTACGTCATC TGTCCTGTCCTGCAGGC (SEQ ID NO: 84). The cDNA was amplified through two successive rounds of PCR using the following conditions: heat denaturation at 95°C for 3 minutes; 35 cycles of a 30 second denaturation (95°C), 30 seconds of annealing (60°C), and 1 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. BIO-X-ACT (BIOLINE) was used to amplify the erythropoietin cDNA from 2.5ng of the human cDNA library in the first round of amplification. Five μ l of the first amplification product was used

-127-

as template for the second round of amplification. Two PCR products were produced from the second amplification with Taq polymerase (Eppendorf), each product was cloned into pCR2.1-Topo (Invitrogen) and sequenced. The larger PCR product contained the expected cDNA
5 sequence for erythropoietin. The erythropoietin cDNA was moved from pTopoEPO into p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110). pTopoEPO was digested with BsiWI and XbaI to release a 588 bp EPO cDNA. BsrGI and BsiWI create compatible ends. The eGFP gene was removed from p18attBZeo(6XHS4)2eGFP by digestion with BsiWI and XbaI, the 8.3 Kbp
10 vector backbone was gel purified and ligated to the 588 bp EPO cDNA to create p18EPOcDNA (SEQ ID NO: 124).

2. Genomic erythropoietin vector, p18genEPO.

The erythropoietin genomic clone was PCR amplified from a human genomic library (Clontech) using the following primers: GENEPO3BSI -
15 CGTACGTCATCTGTCCCCT GTCCTGCA (SEQ ID NO: 85); GENEPO 5XBA -TCTAGAATGGGGGT GCACGGTGAGTACT (SEQ ID NO: 86). The reaction conditions for the amplification were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to
20 50°C), and 3 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds annealing (50°C), and 3 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. The erythropoietin genomic PCR product (2147 bp) was gel purified and cloned into pCR2.1Topo to create pTopogenEPO. Sequence
25 analysis revealed 2bp substitutions and insertions in the intronic sequences of the genomic clone of erythropoietin. A partial digest with XbaI and complete digest with BsiWI excised the erythropoietin genomic insert from pTopogenEPO. The resulting 2158 bp genomic erythropoietin fragment was ligated into the 8.3 Kbp fragment resulting from the

-128-

digestion of p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110) with XbaI and BsrGI to create p18genEPO (SEQ ID NO: 125).

B. TRANSFECTION AND SELECTION WITH DRUG

The erythropoietin genomic and cDNA genes were each moved
5 onto the platform ACes B19-38 (constructed as set forth in Example 3) by
co-transfecting with pCXLamIntR. Control transfections were also
performed using pCXLamInt (SEQ ID NO: 127) together with either
p18EPOcDNA (SEQ ID NO: 124) or p18genEPO (SEQ ID NO: 125).
Lipofectamine Plus was used to transfect the DNA's into B19-38 cells
10 according to the manufacturer's protocol. The cells were placed in
selective medium (DMEM with 10% FBS and Zeocin @ 500ug/ml) 48
hours post-transfection and maintained in selective medium for 13 days.
Clones were isolated 15 days post-transfection.

C. ANALYSIS OF CLONES (ELISA, PCR)

15 1. ELISA Assays

Thirty clones were tested for erythropoietin production by an ELISA
assay using a monoclonal anti-human erythropoietin antibody (R&D
Systems, Catalogue # MAB287), a polyclonal anti-human erythropoietin
antibody (R & D Systems, Catalogue # AB-286-NA) and alkaline
20 phosphatase conjugated goat-anti-rabbit IgG (heavy and light chains)
(Jackson ImmunoResearch Laboratories, Inc., Catalogue # 111-055-144).
The negative control was a Zeocin resistant clone isolated from B19-38
cells transfected with p18attBZeo(6XHS4) (SEQ ID NO: 117; no insert
control vector) and pCXLamIntR (SEQ ID NO: 112). The preliminary
25 ELISA assay was executed as follows: 1) Nunc-Immuno Plates (MaxiSorb
96-well, Catalogue # 439454) were coated with 75ul of a 1/200 dilution
(in Phosphate buffered Saline, pH 7.4 (PBS), Sigma Catalogue # P-3813)
of monoclonal anti-human erythropoietin antibody overnight at 4°C. 2)
The following day the plates were washed 3 times with 300ul PBS

-129-

containing 0.15% Tween 20 (Sigma, Catalogue # P-9416). 3) The plates were then blocked with 300ul of 1% Bovine Serum Albumin (BSA; Sigma Catalogue # A-7030) in PBS for 1 hour at 37°C. 4) Repeat the washes as in step 2. 5) The clonal supernatants (75ul per clone per well of 96-well plate) were then added to the plate and incubated for 1 hour at 37°C. The clonal supernatant analyzed in the ELISA assay had been maintained on the cells 7 days prior to analysis. 6) Repeat the washes of step 2. 7) Add 75ul of polyclonal anti-human erythropoietin antibody (1/250 dilution in dilution buffer (0.5% BSA, 0.01% Tween 20, 1X PBS, pH 7.4) and incubate 1 hour at 37°C. 8) Repeat washes of step 2. 9) Add 75ul of goat anti-rabbit conjugated alkaline phosphatase diluted 1/4000 in dilution buffer and incubate 1 hour at 37°C. 10) Repeat washes of step 2. 11) Add 75ul substrate, p-nitrophenyl phosphate (Sigma N2640), diluted to 1mg/ml in substrate buffer (0.1 Ethanolamine-HCl (Sigma, Catalogue # E-6133), 5mM MgCl₂ (Sigma, Catalogue # M-2393), pH 9.8). Incubate the plates in the dark for 1 hour at room temperature (22°C). 12) Read the absorption at 405nm (reference wavelength 495nm) on an Universal Microplate Reader (Bio-Tek Instruments, Inc., model # ELX800 UV). The erythropoietin standard curve was derived from readings of diluted human recombinant Erythropoietin (Roche, catalogue # 1-120-166; dilution range 125 - 7.8mUnits/ml). From this preliminary assay the 21 clones displaying the highest expression of erythropoietin were analyzed a second time in the same manner using medium supernatants that had been on the clones for 24 hours and a 1:3 dilution thereof.

25 2. PCR Analysis

Genomic DNA was isolated from the 21 clones with the best expression (as assessed by the initial ELISA assay above) as well as the B19-38 cell line and used for PCR analysis. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) according to the

-130-

manufacturers protocol. Amplification was performed on 100ng of genomic DNA as template with MasterTaq DNA Polymerase (Eppendorf) and the primer set 5PacSV40 – CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO: 76) and Antisense Zeo -

- 5 TGAACAGGGTCACGTCGTCC (SEQ ID NO: 77). The amplification conditions were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to 50°C), and 1 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds
- 10 annealing (50°C), and 1 minute extension (72°C); the last cycle is followed by a 10 minute extension at 72°C. PCR products were size separated by gel electrophoresis. Of the 21 clones analyzed 19 produced a PCR product of 650 bp as expected for a site-specific integration event. All nineteen clones were the result of transformations with p19EPOcDNA
- 15 (5) or p18genEPO (14) and pCXLamIntR (i.e. mutant integrase). The remaining two clones, both of which were the result of transformation with p18genEPO (SEQ ID NO: 125) and pCXLamInt (i.e. wildtype integrase; SEQ ID NO: 127), produced a 400 bp PCR product.

Example 12

20 Preparation of a Transformation Vector Useful for the Induction of Plant Artificial Chromosome Formation

- Plant artificial chromosomes (PACs) can be generated by introducing nucleic acid, such as DNA, which can include a targeting DNA, for example rDNA or lambda DNA, into a plant cell, allowing the cell
- 25 to grow, and then identifying from among the resulting cells those that include a chromosome with a structure that is distinct from that of any chromosome that existed in the cell prior to introduction of the nucleic acid. The structure of a PAC reflects amplification of chromosomal DNA, for example, segmented, repeat region-containing and heterochromatic

-131-

structures. It is also possible to select cells that contain structures that are precursors to PACs, for example, chromosomes containing more than one centromere and/or fragments thereof, and culture and/or manipulate them to ultimately generate a PAC within the cell.

5 In the method of generating PACs, the nucleic acid can be introduced into a variety of plant cells. The nucleic acid can include targeting DNA and/or a plant expressable DNA encoding one or multiple selectable markers (*e.g.*, DNA encoding bialophos (bar) resistance) or scorable markers (*e.g.*, DNA encoding GFP). Examples of targeting DNA
10 include, but are not limited to, *N. tabacum* rDNA intergenic spacer sequence (IGS) and *Arabidopsis* rDNA such as the 18S, 5.8S, 26S rDNA and/or the intergenic spacer sequence. The DNA can be introduced using a variety of methods, including, but not limited to *Agrobacterium*-mediated methods, PEG-mediated DNA uptake and electroporation using,
15 for example, standard procedures according to Hartmann *et al* [(1998) *Plant Molecular Biology* 36:741]. The cell into which such DNA is introduced can be grown under selective conditions and can initially be grown under non-selective conditions and then transferred to selective media. The cells or protoplasts can be placed on plates containing a
20 selection agent to grow, for example, individual calli. Resistant calli can be scored for scorable marker expression. Metaphase spreads of resistance cultures can be prepared, and the metaphase chromosomes examined by FISH analysis using specific probes in order to detect amplification of regions of the chromosomes. Cells that have artificial
25 chromosomes with functioning centromeres or artificial chromosomal intermediate structures, including, but not limited to, dicentric chromosomes, formerly dicentric chromosomes, minichromosomes, heterochromatin structures (*e.g.* sausage chromosomes), and stable self-replicating artificial chromosomal intermediates as described herein, are

-132-

identified and cultured. In particular, the cells containing self-replicating artificial chromosomes are identified.

The DNA introduced into a plant cell for the generation of PACs can be in any form, including in the form of a vector. An exemplary
5 vector for use in methods of generating PACs can be prepared as follows.

For the production of artificial chromosomes, plant transformation vectors, as exemplified by pAgIIa and pAgIIb, containing a selectable marker, a targeting sequence, and a scorable marker were constructed using procedures well known in the art to combine the various fragments.
10 The vectors can be prepared using vector pAg1 as a base vector and inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase under the control of the nopaline synthase (NOS) promoter fragment and flanked at the 3' end by the NOS terminator fragment, a fragment of mouse satellite DNA and an *N. tabacum* rDNA intergenic
15 spacer sequence (IGS). In constructing plant transformation vectors, vector pAg2 can also be used as the base vector.

1. Construction of pAG1

Vector pAg1 (SEQ. ID. NO: 89) is a derivative of the CAMBIA vector named pCambia 3300 (Center for the Application of Molecular
20 Biology to International Agriculture, i.e., CAMBIA, Canberra, Australia; www.cambia.org), which is a modified version of vector pCambia 1300 to which has been added DNA from the bar gene conferring resistance to phosphinothricin. The nucleotide sequence of pCambia 3300 is provided in SEQ. ID. NO: 90. pCambia 3300 also contains a lacZ alpha sequence
25 containing a polylinker region.

pAg1 was constructed by inserting two new functional DNA fragments into the polylinker of pCambia 3300: one sequence containing an attB site and a promoterless zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal sequence, and a second

-133-

sequence containing DNA from the hygromycin resistance gene (hygromycin phosphotransferase) conferring resistance to hygromycin for selection in plants. Although the zeomycin-SV40 polyA signal fusion is not expected to function in plant cells, it can be activated in mammalian
5 cells by insertion of a functional promoter element into the attB site by site-specific recombination catalyzed by the Lambda att integrase. Thus, the inclusion of the attB-zeomycin sequences allows for evaluation of functionality of plant artificial chromosomes in mammalian cells by activation of the zeomycin resistance-encoding DNA, and provides an att
10 site for further insertion of new DNA sequences into plant artificial chromosomes formed as a result of using pAg1 for plant transformation. The second functional DNA fragment allows for selection of plant cells with hygromycin. Thus, pAg1 contains DNA from the bar gene conferring resistance to phosphinothricin, DNA from the hygromycin resistance gene,
15 both resistance-encoding DNAs under the control of a separate cauliflower mosaic virus (CaMV) 35S promoter, and the attB-promoterless zeomycin resistance-encoding DNA.

pAg1 is a binary vector containing *Agrobacterium* right and left T-DNA border sequences for use in *Agrobacterium*-mediated transformation
20 of plant cells or protoplasts with the DNA located between the border sequences. pAg1 also contains the pBR322 Ori for replication in *E. coli*. pAg1 was constructed by ligating *HindIII*/*PstI*-digested p3300attBZeo with *HindIII*/*PstI*-digested pBSCaMV35SHyg as follows.

a. Generation of p3300attBZeo

25 Plasmid pCambia 3300 was digested with *PstI*/*Eco*136 II and ligated with *PstI*/*StuI*-digested pLITattBZeo (the nucleotide sequence of pLITattBZeo is provided in SEQ. ID. NO: 91. (containing DNA encoding the zeocin resistance gene and an attB Integrase recognition sequence) to generate p3300attBZeo which contains an attB site, a promoterless

-134-

zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal, and a reconstructed *Pst*I site.

b. Generation of pBSCaMV35SHyg

A DNA fragment containing DNA encoding hygromycin

5 phosphotransferase flanked by the CaMV 35S promoter and the CaMV 35S polyA signal sequence was obtained by PCR amplification of plasmid pCambia 1302 (GenBank Accession No. AF234298 and SEQ. ID. NO: 92). The primers used in the amplification reaction were as follows: CaMV35SpolyA:

10 5'-CTGAATTAACGCCGAATTAATTCGGGGGATCTG-3' SEQ. ID. NO: 93
CaMV35Spr:

5'-CTAGAGCAGCTTGCCAACATGGTGGAGCA-3' SEQ. ID. NO: 94

The 2100-bp PCR fragment was ligated with *Eco*RV-digested pBluescript II SK+ (Stratagene, La Jolla, CA, U.S.A.) to generate pBSCaMV35SHyg.

15 **c. Generation of pAg1**

To generate pAg1, pBSCaMV35SHyg was digested with *Hind*III/*Pst*I and ligated with *Hind*III/*Pst*I-digested p3300attBZeo. Thus, pAg1 contains the pCambia 3300 backbone with DNA conferring resistance to phosphinothricin and hygromycin under the control of
20 separate CaMV 35S promoters, an attB-promoterless zeomycin resistance-encoding DNA recombination cassette and unique sites for adding additional markers, *e.g.*, DNA encoding GFP. The attB site can be used as described hereing for the addition of new DNA sequences to plant artificial chromosomes, including PACs formed as a result of using the
25 pAg1 vector, or derivatives thereof, in the production of PACs. The attB site provides a convenient site for recombinase-mediated insertion of DNAs containing a homologous att site.

2. pAG2

-135-

The vector pAg2 (SEQ. ID. NO: 95) is a derivative of vector pAg1 formed by adding DNA encoding a green fluorescent protein (GFP), under the control of a NOS promoter and flanked at the 3' end by a NOS polyA signal, to pAg1. pAg2 was constructed as follows. A DNA fragment
5 containing the NOS promoter was obtained by digestion of pGEM-T-NOS, or pGEMEasyNOS (SEQ. ID. NO: 96), containing the NOS promoter in the cloning vector pGEM-T-Easy (Promega Biotech, Madison, WI, U.S.A.), with *Xba*I/*Nco*I and was ligated to an *Xba*I/*Nco*I fragment of pCambia 1302 containing DNA encoding GFP (without the CaMV 35S promoter) to
10 generate p1302NOS (SEQ. ID. NO: 97) containing GFP-encoding DNA in operable association with the NOS promoter. Plasmid p1302NOS was digested with *Sma*I/*Bs*WI to yield a fragment containing the NOS promoter and GFP-encoding DNA. The fragment was ligated with *Pme*I/*Bs*WI-digested pAg1 to generate pAg2. Thus, pAg2 contains DNA
15 from the bar gene conferring resistance to phosphinothricin, DNA conferring resistance to hygromycin, both resistance-encoding DNAs under the control of a cauliflower mosaic virus 35S promoter, DNA encoding kanamycin resistance, a GFP gene under the control of a NOS promoter and the attB-zeomycin resistance-encoding DNA. One of skill in
20 the art will appreciate that other fragments can be used to generate the pAg1 and pAg2 derivatives and that other heterologous DNA can be incorporated into pAg1 and pAg2 derivatives using methods well known in the art.

3. pAgIIa and pAgIIb transformation vectors

25 Vectors pAgIIa and pAgIIb were constructed by inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase, the nopaline synthase terminator fragment, the nopaline synthase (NOS) promoter fragment, a fragment of mouse satellite DNA and an *N. tabacum*

-136-

rDNA intergenic spacer sequence (IGS). The construction of pAgIIa and pAgIIb was as follows.

An *N. tabacum* rDNA intergenic spacer (IGS) sequence (SEQ. ID. NO: 98; see also GenBank Accession No. YO8422; see also Borysyuk *et al.* (2000) *Nature Biotechnology* 18:1303-1306; Borysyuk *et al.* (1997) *Plant Mol. Biol.* 35:655-660; U.S. Patent Nos. 6,100,092 and 6,355,860) was obtained by PCR amplification of tobacco genomic DNA. The IGS can be used as a targeting sequence by virtue of its homology to tobacco rDNA genes; the sequence is also an amplification promoter sequence in plants. This fragment was amplified using standard PCR conditions (*e.g.*, as described by Promega Biotech, Madison, WI, U.S.A.) from tobacco genomic DNA using the primers shown below:

NTIGS-FI

5'- GTG CTA GCC AAT GTT TAA CAA GAT G- 3' (SEQ ID No. 99) and

NTIGS-RI

5'-ATG TCT TAA AAA AAA AAA CCC AAG TGA C- 3' (SEQ ID No. 100)

Following amplification, the fragment was cloned into pGEM-T Easy to give pIGS-I. A fragment of mouse satellite DNA (Msat1 fragment; GenBank Accession No. V00846; and SEQ ID No. 101) was amplified via PCR from pSAT-1 using the following primers:

MSAT-F1

5'- AAT ACC GCG GAA GCT TGA CCT GGA ATA TCG C -3' (SEQ ID No. 102) and

MSAT-RI

5'-ATA ACC GCG GAG TCC TTC AGT GTG CA T- 3' (SEQ ID No. 103)

This amplification added a *SacII* and a *HindIII* site at the 5' end and a *SacII* site at the 3' end of the PCR fragment. This fragment was then cloned into the *SacII* site in pIGS-1 to give pMIGS-1, providing a eukaryotic

-137-

centromere-specific DNA and a convenient DNA sequence for detection via FISH.

A functional marker gene containing a NOS-promoter:GUS:NOS terminator fusion was then constructed containing the NOS promoter
5 (GenBank Accession No. U09365; SEQ ID No. 104), *E. coli*
 β -glucuronidase coding sequence (from the GUS gene; GenBank
Accession No. S69414; and SEQ ID No. 105), and the nopaline synthase
terminator sequence (GenBank Accession No. U09365; SEQ ID No. 107).
The NOS promoter in pGEM-T-NOS was added to a promoterless GUS
10 gene in pBlueScript (Stratagene, La Jolla, CA, U.S.A.) using *NotI*/*SpeI* to
form pNGN-1, which has the NOS promoter in the opposite orientation
relative to the GUS gene.

pMIGS-1 was digested with *NotI*/*SpeI* to yield a fragment
containing the mouse major satellite DNA and the tobacco IGS which was
15 then added to *NotI*-digested pNGN-1 to yield pNGN-2. The NOS promoter
was then re-oriented to provide a functional GUS gene, yielding pNGN-3,
by digestion and religation with *SpeI*. Plasmid pNGN-3 was then digested
with *HindIII*, and the *HindIII* fragment containing the β -glucuronidase
coding sequence and the rDNA intergenic spacer, along with the Msat
20 sequence, was added to pAG-1 to form pAgIIa (SEQ ID NO: 108), using
the unique *HindIII* site in pAg1 located near the right T-DNA border of
pAg1, within the T-DNA region.

Another plasmid vector, referred to as pAgIIb, was also recovered,
which contained the inserted *HindIII* fragment (SEQ ID NO: 108) in the
25 opposite orientation relative to that observed in pAgIIa. Thus, pAgIIa and
pAgIIb differ only in the orientation of the *HindIII* fragment containing the
mouse major satellite sequence, the GUS DNA sequence and the IGS
sequence. The nucleotide sequences of pAgIIa is provided in SEQ. ID.
NOS: 109.

-138-

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

-139-

WHAT IS CLAIMED IS:

1. A eukaryotic chromosome comprising one or a plurality of *att* site(s), wherein:
 - an *att* site is heterologous to the chromosome; and
 - 5 an *att* site permits site-directed integration in the presence of lambda integrase.
2. The eukaryotic chromosome of claim 1, wherein the *att* sites are selected from the group consisting of *attP* and *attB* or *attL* and *attR*, or variants thereof.
- 10 3. The eukaryotic chromosome of claim 1 that is an artificial chromosome.
4. The eukaryotic chromosome of claim 1 that is an artificial chromosome expression system (*ACes*).
5. The eukaryotic chromosome of claim 4 that is predominantly
15 heterochromatin.
6. The chromosome of claim 1 that is an artificial chromosome that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
7. The chromosome of claim 1 that is a plant chromosome.
- 20 8. The chromosome of claim 1 that is an animal chromosome.
9. The chromosome of claim 7 that is a plant artificial chromosome.
10. The chromosome of claim 8 that is an animal artificial chromosome.
- 25 11. The chromosome of claim 8 that is a mammalian chromosome.
12. The chromosome of claim 11 that is a mammalian artificial chromosome.

-140-

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).

14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase
5 catalyzed recombination.

15. The *ACes* of claim 14 that contains one site.

16. The *ACes* of claim 14 that is predominantly heterochromatin.

17. The *ACes* of claim 14 that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.

10 18. The *ACes* of claim 14 that is a plant *ACes*.

19. The *ACes* of claim 14 that is an animal *ACes*.

20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.

21. The *ACes* of claim 14 that is a fish *ACes*.

15 22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli* and the R/RS
20 system of the pSR1 plasmid, or any combination thereof.

23. A method of introducing heterologous nucleic acid into a chromosome, comprising:

contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a
25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

-141-

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.
25. The method of claim 23, wherein the nucleic acid molecule
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.
26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).
- 10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.
28. The combination of claim 27, further comprising nucleic acid
15 encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.
29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase
20 system of phage Mu, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.
30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.
31. The combination of claim 27, wherein a vector is the plasmid
25 pDsRedN1-attB.
32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

-142-

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

- (a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second recombination site and the heterologous nucleic acid;
- (b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis and Cre, $\gamma\delta$, Tn3 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 32, wherein the recombination sites are selected from the group consisting of att and lox P sites.

-143-

41. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that removes one or more stop codons.

42. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that avoids hairpin formation.

43. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence selected from the group consisting of SEQ ID NOs:41-56:

- 10 a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:41);
- b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:42);
- c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:43);
- d) AGCCWGCTTTCKTRTACNAAAGTSGB (m-attL) (SEQ ID NO:44);
- e) GTTCAGCTTTYKTRTACNAAAGTSGB (m-attP1) (SEQ ID NO:45);
- 15 f) AGCCTGCTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:46);
- g) AGCCTGCTTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:47);
- h) ACCCAGCTTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:48);
- i) GTTCAGCTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:49);
- j) GTTCAGCTTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:50);
- 20 k) GTTCAGCTTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:51);
- l) AGCCTGCTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:52);
- m) AGCCTGCTTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:53);
- n) ACCCAGCTTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:54);
- o) GTTCAGCTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:55);
- 25 p) GTTCAGCTTTCTTGTACAAAGTTGG (attP2, P3) (SEQ ID NO: 56);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G, K=G or T/U, Y=C or T/U, W=A or T/U, N=A or C or G or T/U, S=C or G, and B=C or G or T/U; and

-144-

the core region does not contain a stop codon in one or more reading frames.

44. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence
5 selected from the group consisting of a mutated att recombination site containing at least one mutation that enhances recombinational specificity, a complementary DNA sequence thereto, and an RNA sequence corresponding thereto.

45. The method of claim 33, wherein the vector comprising the
10 second site further encodes at least one selectable marker.

46. The method of claim 45, wherein the marker is a promoterless marker, which, upon recombination is under the control of a promoter and is thereby expressed.

47. The method of claim 46, wherein the first recombination site
15 is attP and is in the sense orientation prior to recombination.

48. The method of claim 46, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic, fluorescent, or capable of being bound by an antibody and FACs sorted.

20 49. The method of claim 48, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase (hisD).

50. A cell comprising, the chromosome of claim 1.

25 51. The cell of claim 50, wherein the cell is a nuclear donor cell.

52. The cell of claim 50, wherein the cell is a stem cell.

53. The stem cell of claim 52, wherein said stem cell is human and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

-145-

54. The cell of claim 50, wherein the cell is mammalian.
55. The cell of claim 54, wherein the mammal is selected from the group consisting of humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, and horses.
- 5 56. The cell of claim 50, wherein the cell is a plant cell.
57. A cell comprising the platform *ACes* of claim 14.
58. The cell of claim 57, wherein the cell is a nuclear donor cell.
59. The cell of claim 57, wherein the cell is a stem cell.
60. The stem cell of claim 59, wherein said stem cell is human
- 10 and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.
61. A human mesenchymal cell comprising an artificial chromosome.
62. The human mesenchymal cell of claim 61, wherein said
- 15 artificial chromosome is an *ACes*.
63. The human mesenchymal cell of claim 62, wherein the *ACes* is a platform-*ACes*.
64. A method for introducing heterologous nucleic acid into the mesenchymal cell of claim 63, comprising:
- 20 (a) introducing into the cell of claim 63, wherein the platform-*ACes* has a first recombination site, a vector comprising at least a second recombination site and the heterologous nucleic acid;
- (b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between
- 25 the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the platform-*ACes* within the mesenchymal cell.
65. A lambda-intR mutin comprising a glutamic acid to arginine change at position 174 of wild-type lambda-intR.

-146-

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence
10 of a IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or
fluorescent.

15 70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

25 74. The method of claim 72, wherein the embryonic cell is in an embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

-147-

76. The method of claim 72, wherein the transgenic animal is a fish, insect, reptile, amphibians, arachnid or mammal.

77. The method of claim 72, wherein the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection,
5 microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

78. A transgenic animal produced by the method of claim 72.

79. A cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated
10 throughout the endogenous chromosomes.

80. A method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into the cell line of claim 79, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of
15 the heterologous recombination sites within the cell's chromosomal DNA; and isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within.

81. A library of cells useful for genomic screening, said library
20 comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein.

82. The library of cells of claim 81, wherein the cells of the library are from a different species than the chromosomal nucleic acid
25 within the *ACes*.

83. A method of making one or more cell lines, comprising
a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,

-148-

b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;

5 c) isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;

 d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;

10 e) selecting different cells having mutually exclusive *ACes* therein and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

84. The method of claim 23, wherein the nucleic acid molecule with a recombination site is a PCR product.

15 85. Method of claim 23 wherein the recombinase is a protein and the recombination event occurs in vitro.

86. The method of claim 33, wherein the vector is a PCR product comprising a second recombination site.

20 87. The lambda-intR mutein of claim 65, wherein the mutein further comprises an amino acid signal for nuclear localization.

88. The lambda-intR mutein of claim 65, wherein the mutein further comprises an epitope tag for protein purification.

89. A modified iron-induced promoter comprising SEQ ID NO:128.

25 90. A plasmid or expression cassette comprising the promoter of claim 89.

91. A vector, comprising:
a recognition site for recombination; and

-149-

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

-150-

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:

- 5 introducing a vector of claim 91 into a cell;
 growing the cells; and
 selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein sufficient portion of the
10 vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:

- 15 contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

 the reporter *ACes* comprises one or a plurality of reporter constructs;

- a reporter construct comprises a reporter gene in operative linkage
20 with a regulatory region responsive to test or known compounds; and

 detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively
25 linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

-151-

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent
5 proteins.

113. The method of claim 109, wherein the *ACes* comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

-152-

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cell.

122. The method of claim 109, wherein the cell is in an animal;
and the method comprises whole-body imaging to monitor expression of
5 the reporter in the animal.

123. A reporter *ACes* comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

Fig. 1 Generation of ACes for Platform Chromosome Engineering

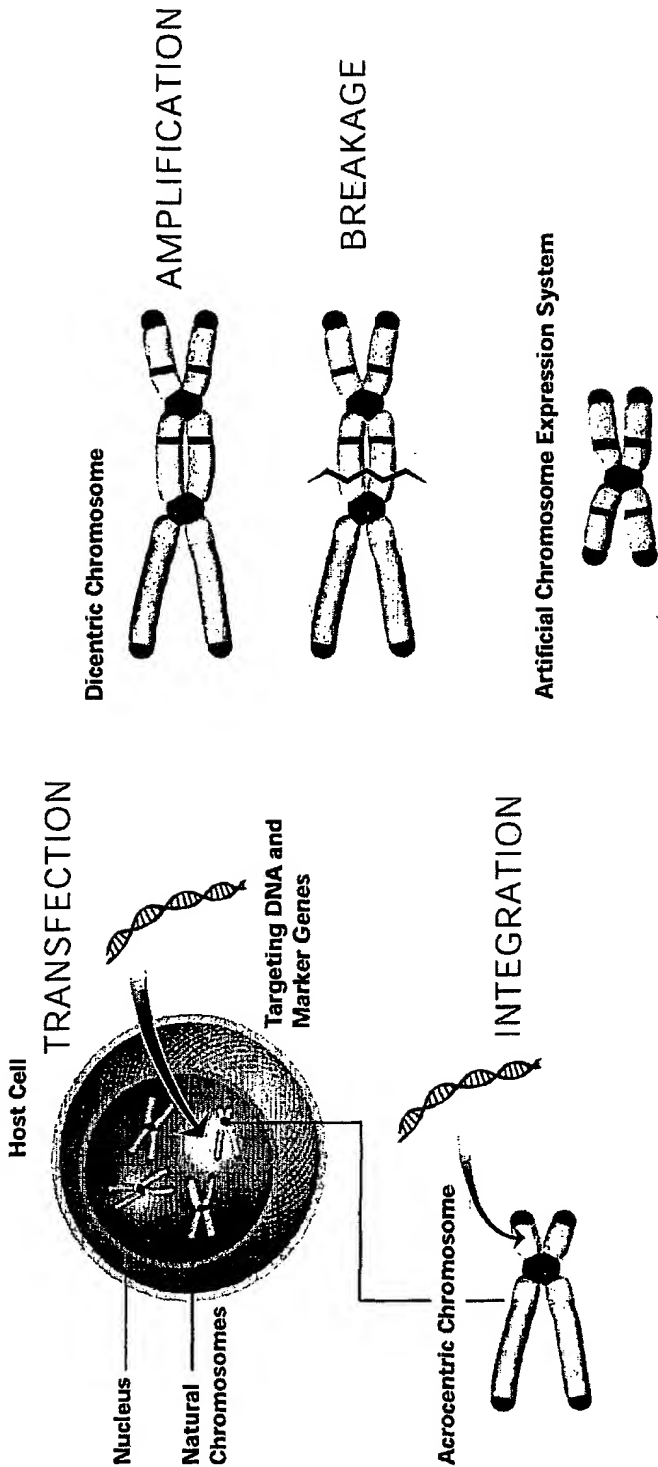


Fig. 2 Targeting plasmid pWEPuro9K

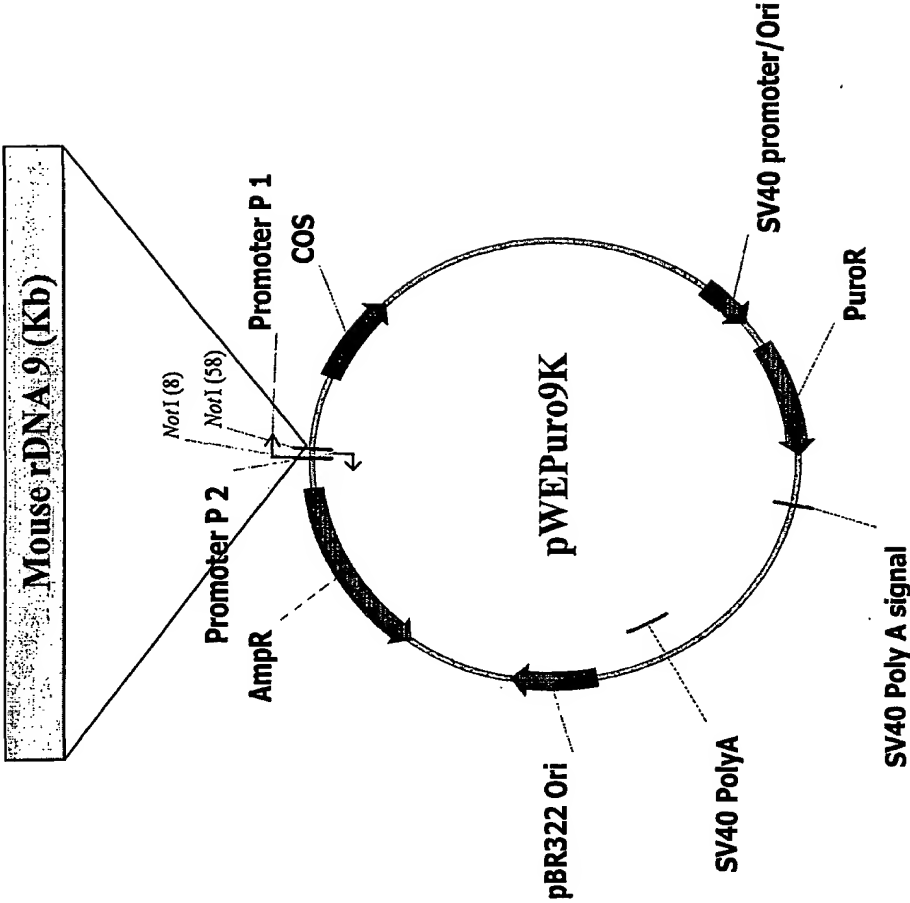


Fig. 3 Platform Chromosome with single recombination site

Platform chromosome containing mouse rDNA, mouse major/minor repeat DNA and puromycin resistant cassette



Targeting onto platform chromosome via homologous recombination



or



Engineered Selectable Marker System

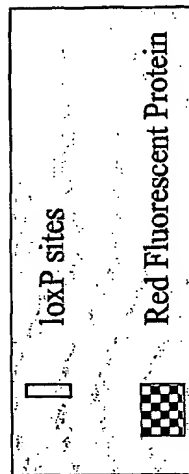
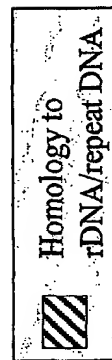
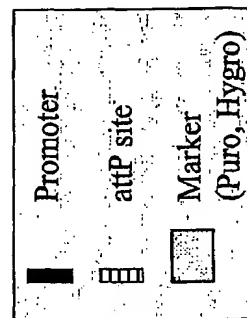


Fig. 4

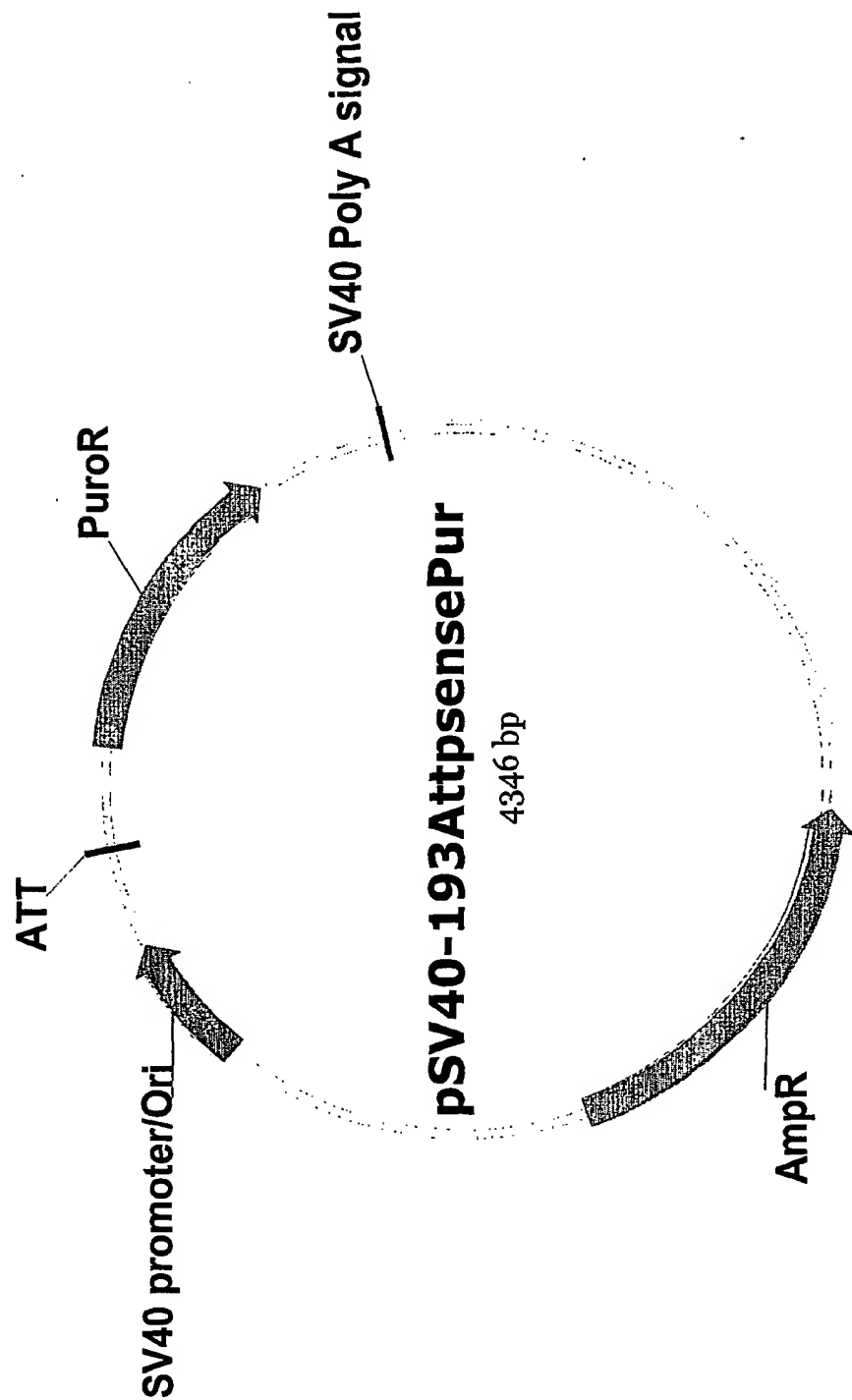


Fig. 5 Generation of chromosome based platform with multiple integration sites

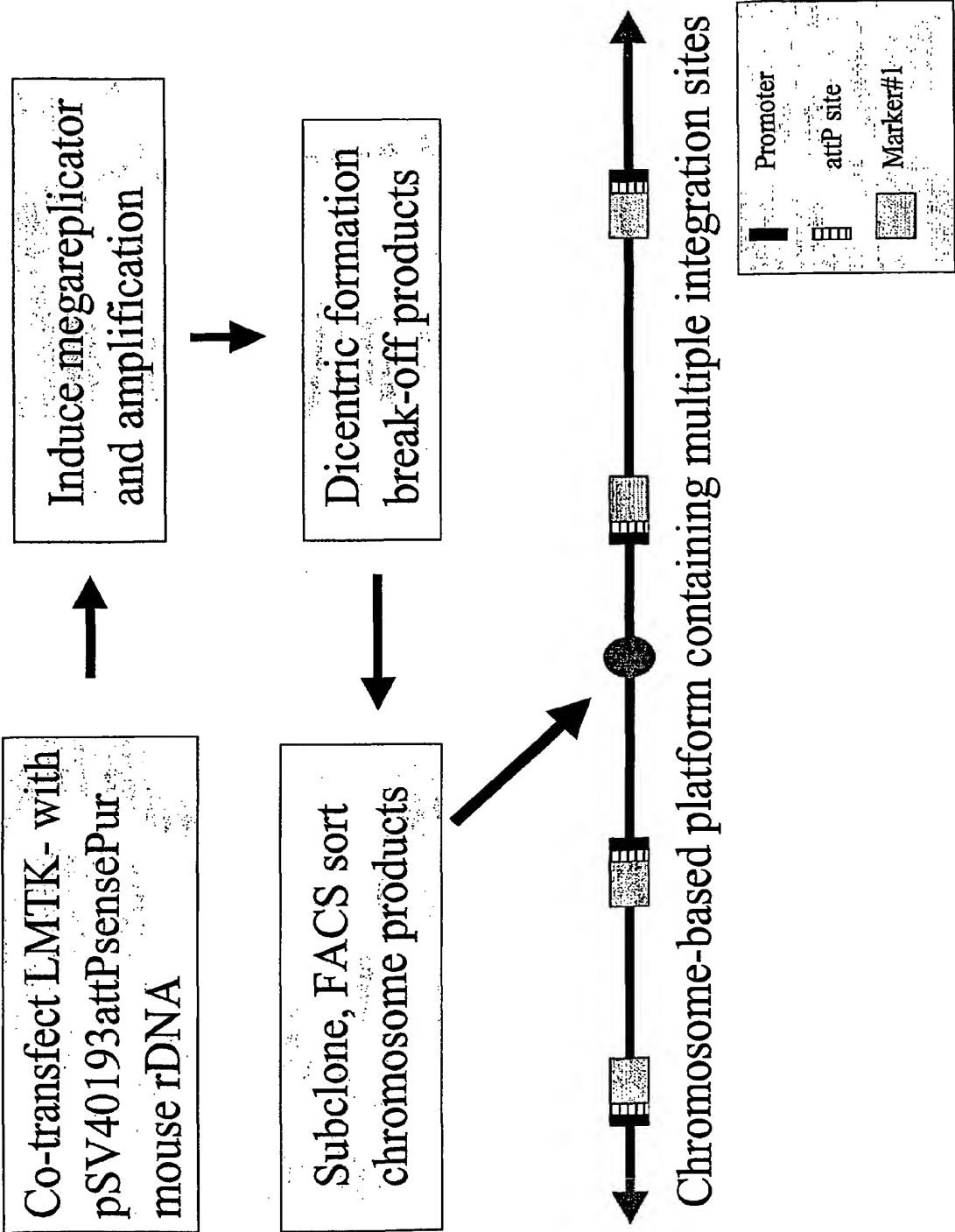


Fig. 6 λ integrase recombination



Core Region	
attP	CAGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attB	CTGCTTTT <u>T</u> TATACTAA <u>C</u> TTG
attL	CTGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attR	CAGCTTTT <u>T</u> TATACTAA <u>C</u> TTG

Fig. 7 λ INT recombination on artificial chromosome

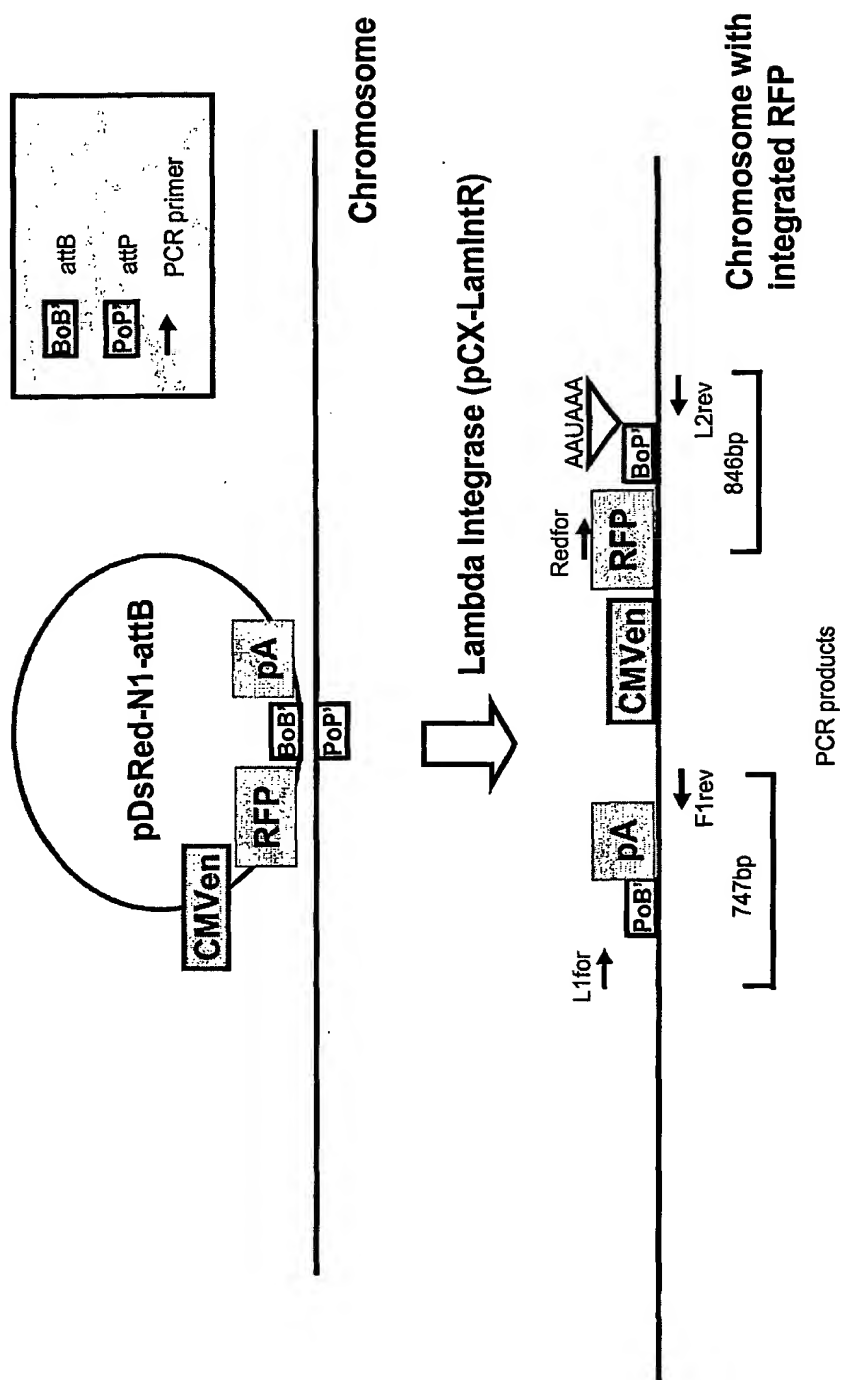
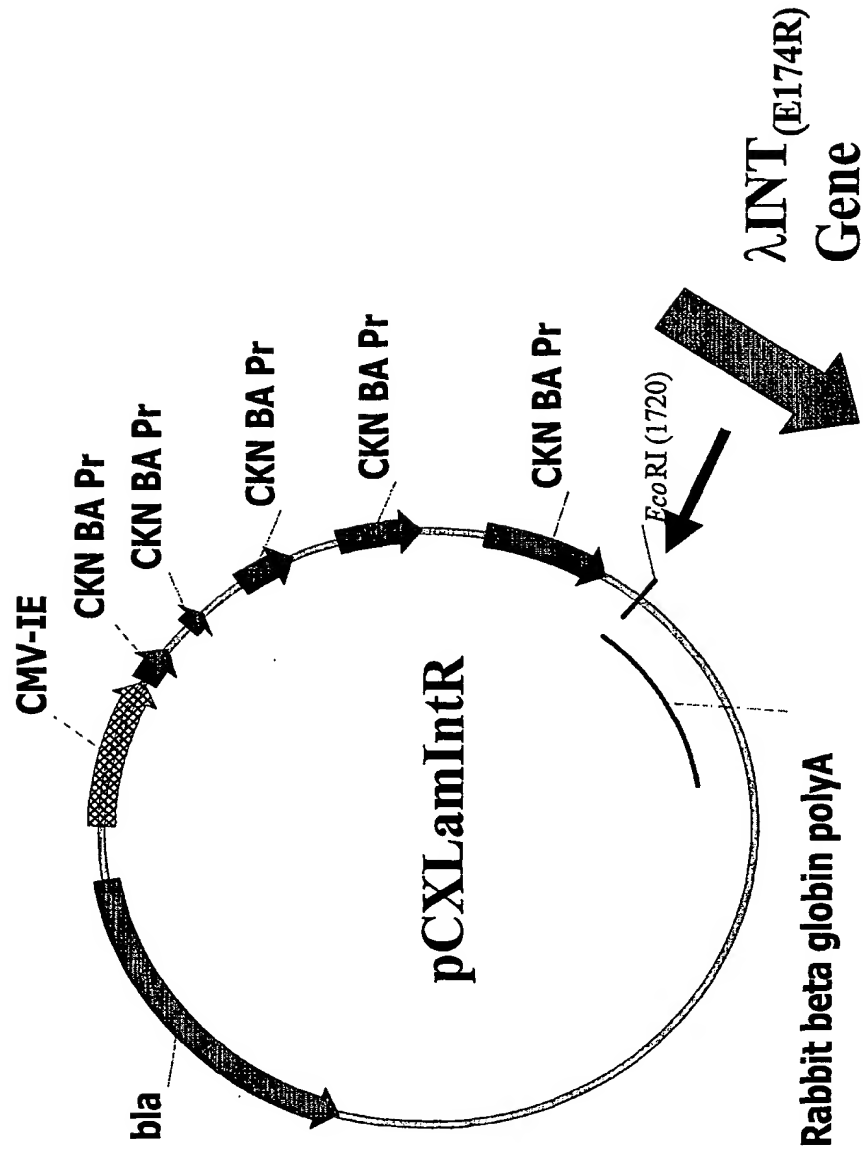


Fig. 8 pCXLaIntR integrase expression vector



9/15

Fig. 9 Platform Technology

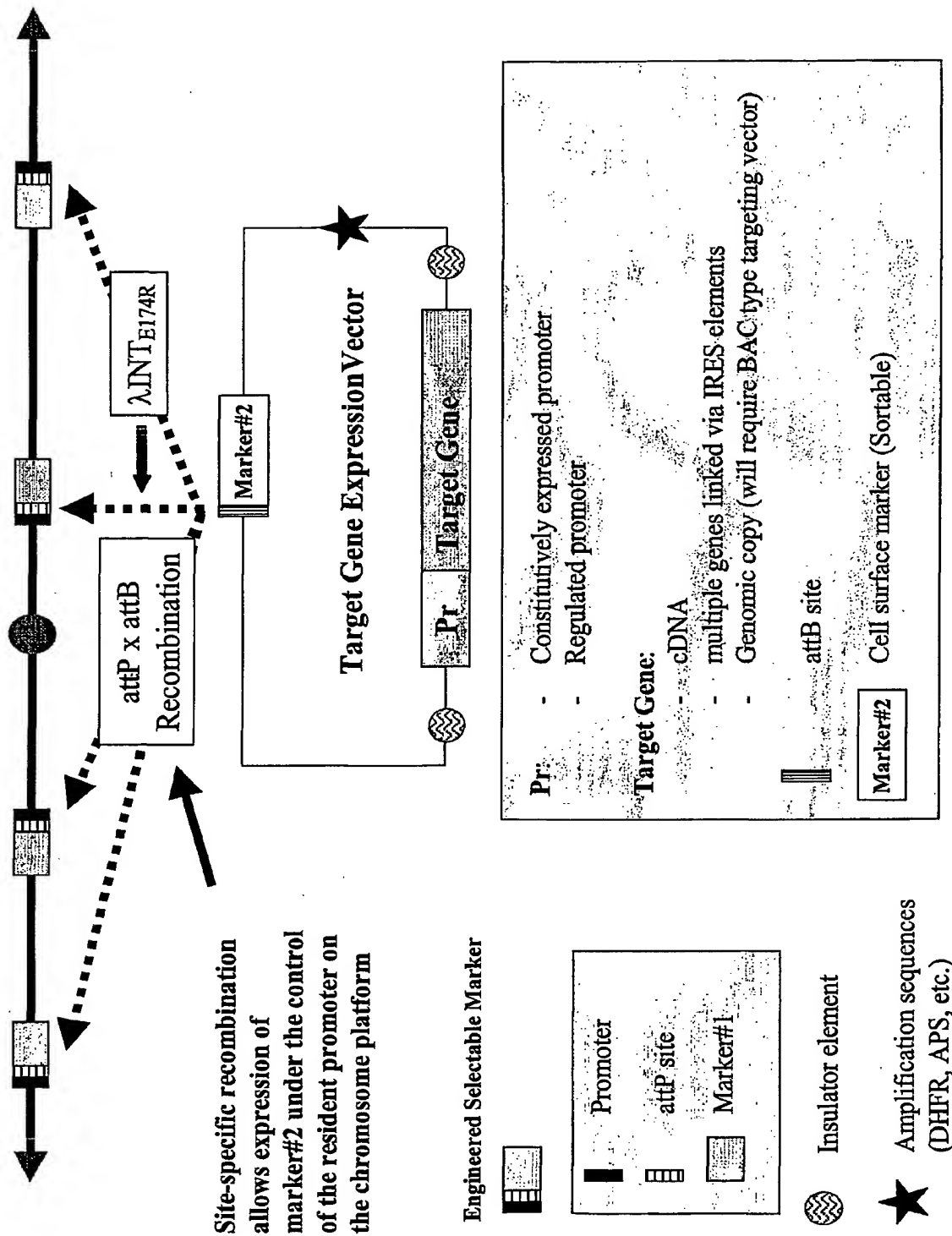


Fig. 10

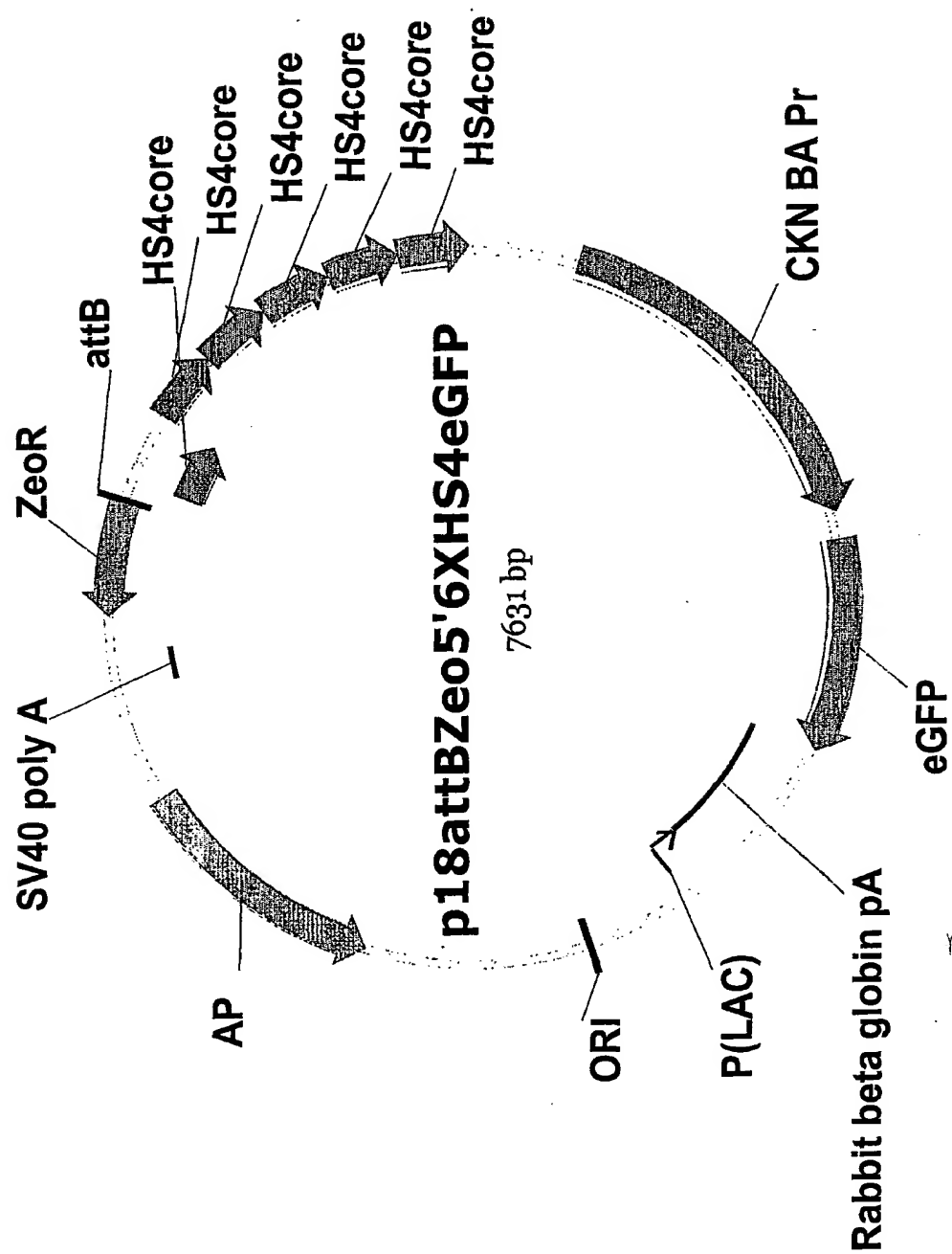
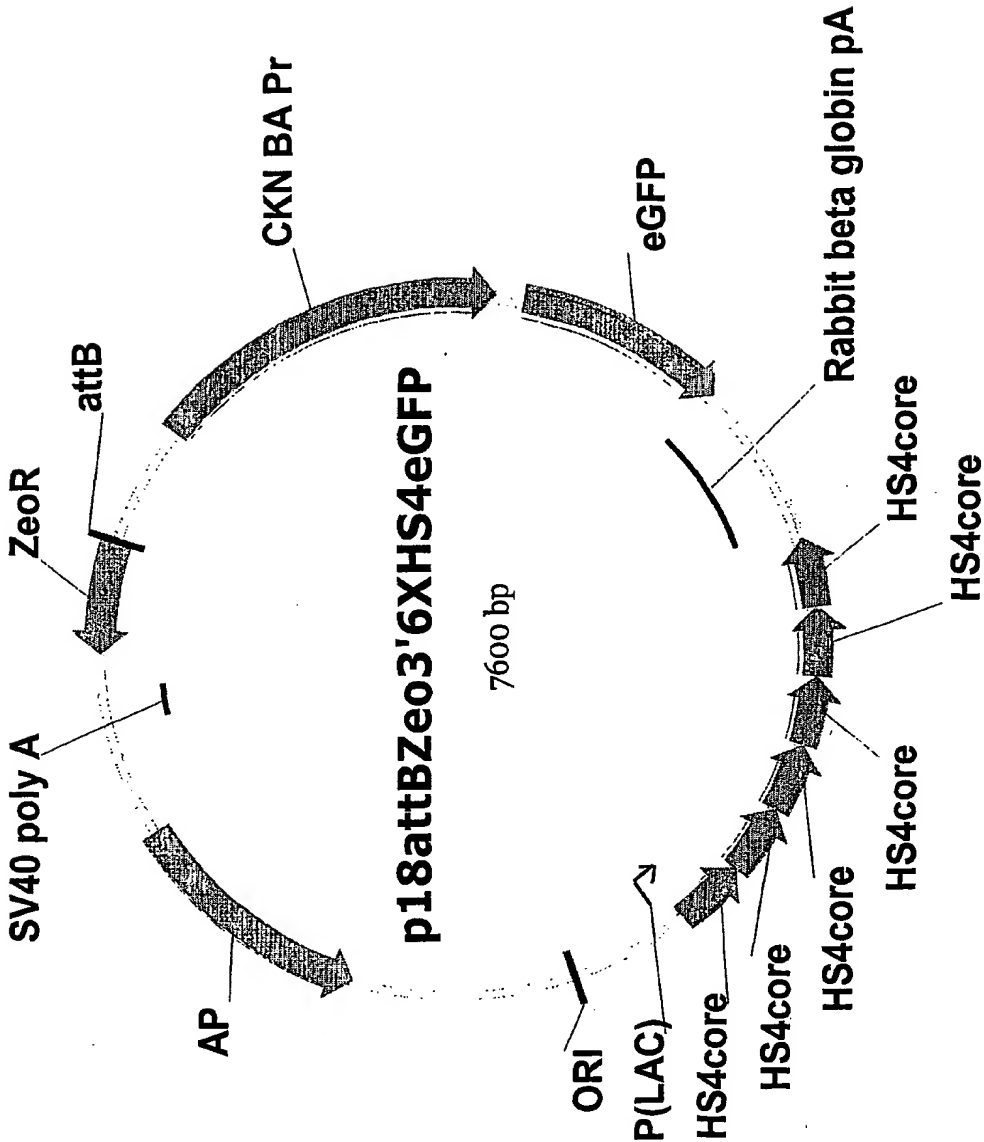


Fig. 11



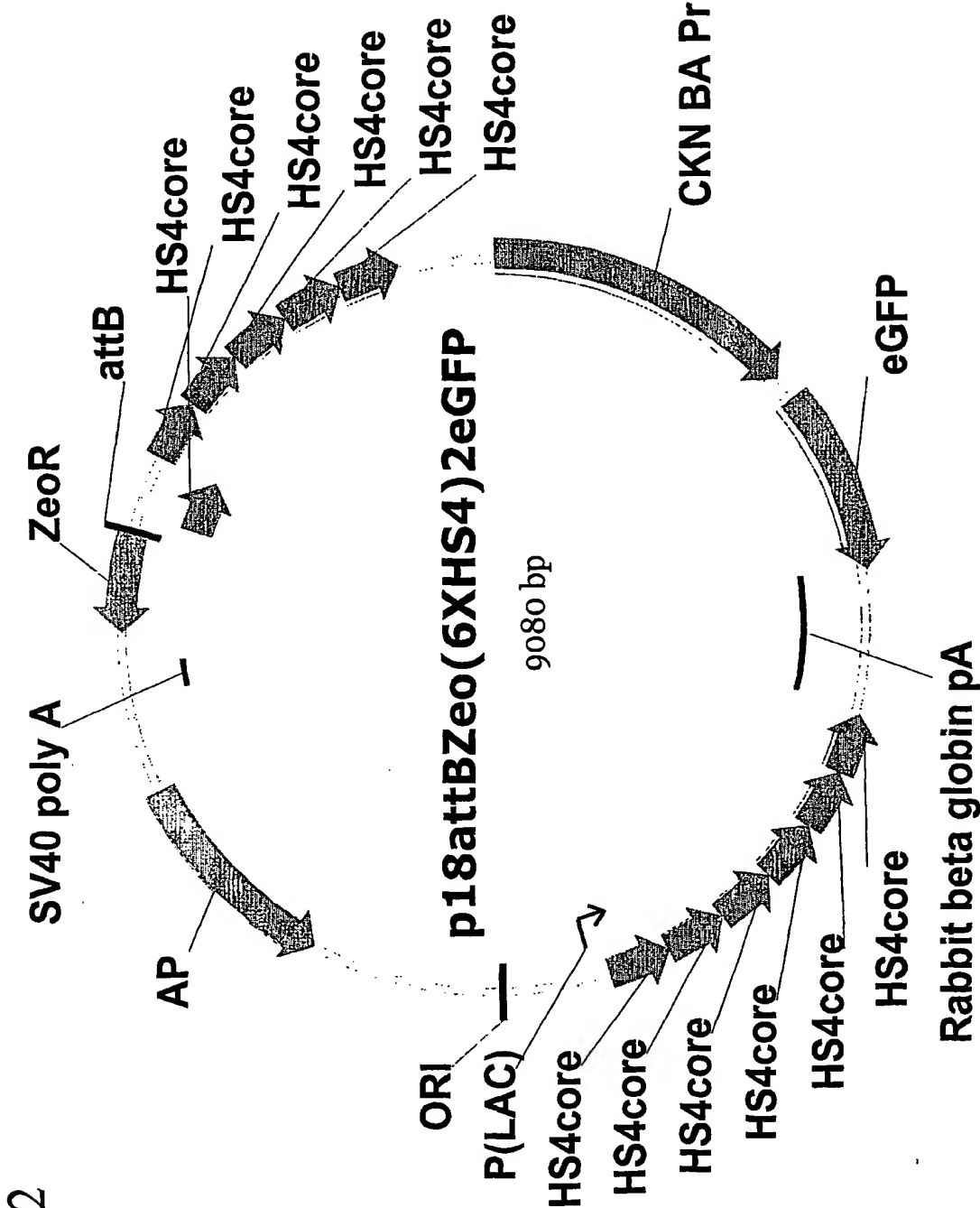
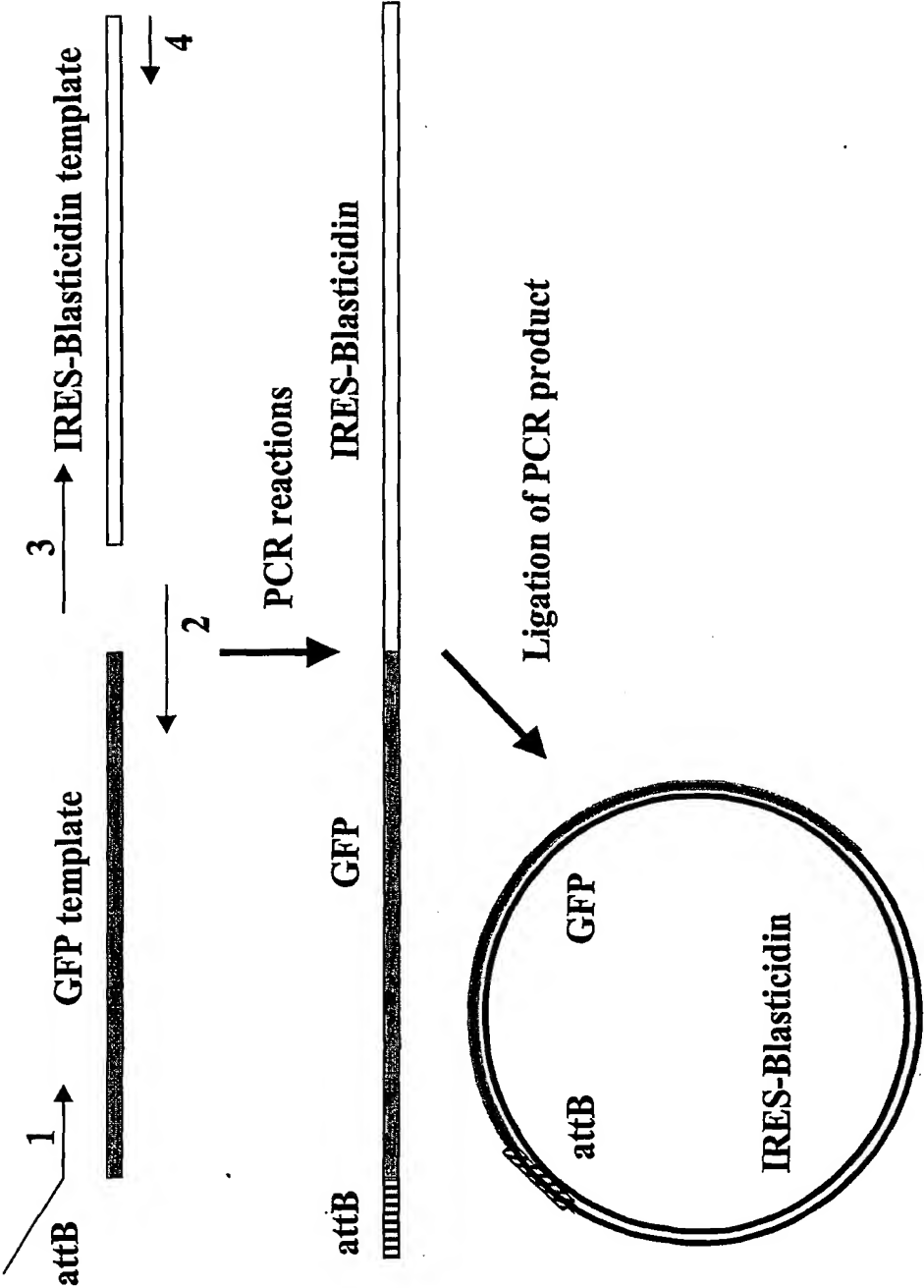


Fig. 12

Fig. 13



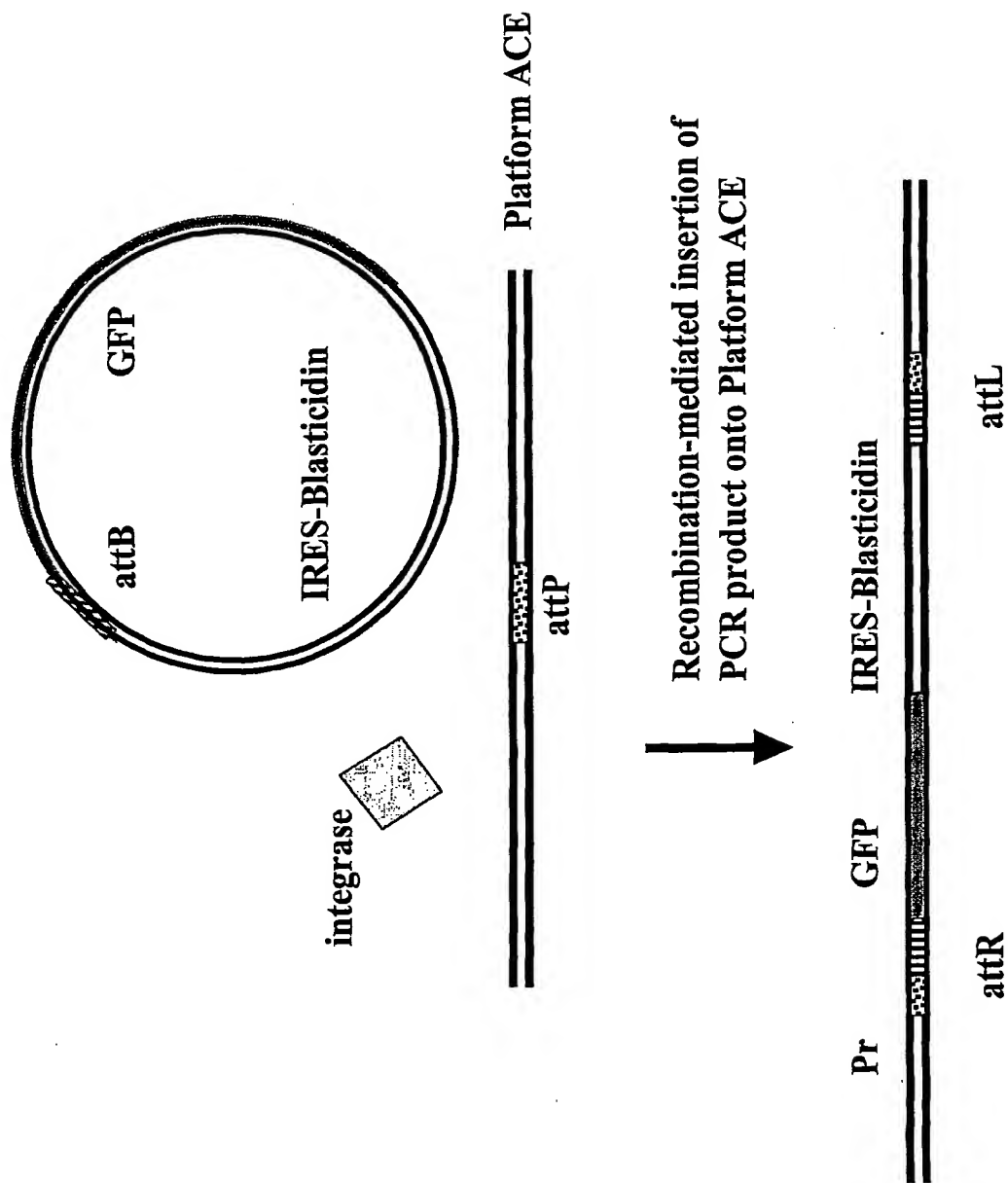


Fig. 14

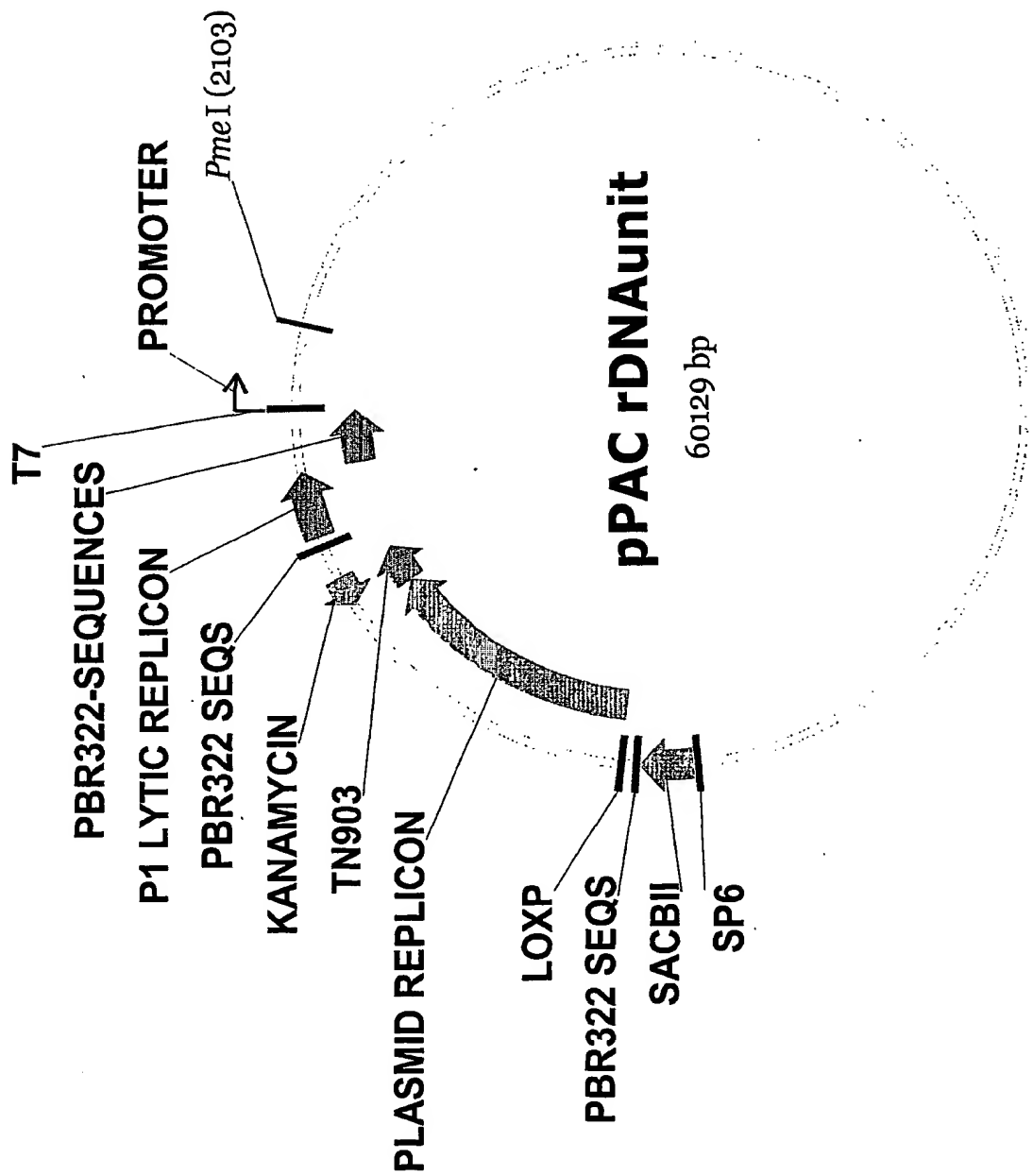


Fig. 15

-1-

SEQUENCE LISTING

<110> CHROMOS MOLECULAR SYSTEMS, INC.
Perkins, Edward
Perez, Carl
Lindenbaum, Michael
Greene, Amy
Leung, Josephine
Fleming, Elena
Stewart, Sandra
Shellard, Joan

<120> CHROMOSOME-BASED PLATFORMS

<130> 24601-420PC

<140> Not Yet Assigned
<141> Herewith

<150> 60/294,758
<151> 2001-05-30

<150> 60/366,891
<151> 2002-03-21

<160> 129

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: attPUP

<400> 1
ccttgcgcta atgctctgtt acagg 25

<210> 2
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: attPDWN

<400> 2
cagaggcagg gagtgggaca aaattg 26

<210> 3
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer: Lamint 1

<400> 3
ttcgaattca tgggaagaag gcgaagtcac gagcg 35

<210> 4
<211> 34
<212> DNA
<213> Artificial Sequence

-2-

<220>
<223> Primer: Lamint 2

<400> 4
ttcgaattct tatttgattt caattttgtc ccac 34

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 5
cggacaatgc ggttgtgcgt 20

<210> 6
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 6
cgcgacagcaa aatctagagt aaggagatca agacttacgg ctgacg 46

<210> 7
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> LambdaINTER174rev

<400> 7
cgtcagccgt aagtcttgat ctccttactc tagattttgc tgcgcg 46

<210> 8
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 8
tgaagcctgc ttttttatac taacttgagc gaa 33

<210> 9
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 9
ttcgctcaag ttagtataaa aaagcaggct tca 33

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

-3-

<223> Primer: attPdwn2

<400> 10

tcttctcggg cataagtcgg acacc

25

<210> 11

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer:CMVen

<400> 11

ctcacgggga tttccaagtc tccac

25

<210> 12

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer:attPdwn

<400> 12

cagaggcagg gagtgggaca aaattg

26

<210> 13

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer:CMVEN2

<400> 13

caactccgcc ccattgacgc aaatg

25

<210> 14

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer:L1

<400> 14

agtatcgccg aacgattagc tcttca

26

<210> 15

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer:F1 rev

<400> 15

gccgatttcg gcctattggt taaa

24

<210> 16

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer:RED

-4-

<400> 16
ccgccgacat ccccgactac aagaa

25

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L2rev

<400> 17
ttccttcgaa ggggatccgc ctacc

25

<210> 18
<211> 22118
<212> DNA
<213> Mus musculus

<300>
<308> GenBank X82564
<309> 1996-04-09

<400> 18
gaattccctt atccctaatt cagattggtg gaataacttg gtatagatgt ttgtgcatta 60
aaaaccctgt aggatcttca ctctaggtca ctgttcagca ctggaacctg aatttgtggcc 120
ctgagtgata ggtcctggga catatgcagt tctgcacaga cagacagaca gacagacaga 180
cagacagaca gacagacgtt acaaacaaac acgttgagcc gtgtgccaac acacacacaa 240
acaccactct ggccataatt attgaggacg ttgatttatt attctgtgtt tgtgagtcctg 300
tctgtctgtc tgtctgtctg tctgtctgtc tatcaaacca aaagaaacca aacaattatg 360
cctgcctgcc tgctgcctg cctacacaga gaaatgattt cttcaatcaa tctaaaacga 420
cctcctaagt ttgccttttt tctctttctt tatctttttc ttttttcttt tcttcttctt 480
tccttccttc ctctccttct tccttccttt cttctttctt ttctttcttt cttactttct 540
ttctttcctt cttacattta ttcttttcat acatagtttc ttagtgtaag catccctgac 600
tgtcttgaag acactttgta ggccctcaat ctgtaagagc ctctctctgc ttttcaaagt 660
ctggcatgaa tgttgtacct cactatgacc agcttagtct tcaagtctga gttactggaa 720
aggagttcca agaagactgg ttatattttt catttattat tgcattttta ttaaaattta 780
atctcaccaa aagaatttag actgaccaat tcagagtctg ccgtttaaaa gcataaggaa 840
aaagtaggag aaaaacgtga ggctgtctgt ggatggtcga ggctgcttta gggagcctcg 900
tcaccattct gcacttgcaa accggggccac tagaaccctg tgaagggaga aaccaaagcg 960
acctggaaac aataggtcac atgaaggcca gccacctcca tcttgttgtg cgggagttca 1020
gttagcagac aagatggctg ccattgcacat gttgtctttc agcttggtga ggtcaaagta 1080
caaccgagtc acagaacaag gaagtataca cagtgaagtc caggtcagcc agagtttaca 1140
cagagaaacc acatcttgaa aaaaacaaaa aaataaatta aataaatata atttaaaaat 1200
ttaaaaaatag ccgggagtg tggcgcatgt ctttaatccc agctctcttc aggcagagat 1260
gggaggattt ctgagtttga ggccagcctg gtctgcaaag tgagttccag gacagtcagg 1320
gctatacaga gaaacctgt cttgaaaact aaactaaatt aaactaaact aaactaaaaa 1380
aatataaaat aaaaatttta aagaatttta aaaaactaca gaaatcaaac ataagcccac 1440
gagatggcaa gtaactgcaa tcatagcaga aatattatac acacacacac acacagactc 1500
tgtcataaaa tccaatgtgc cttcatgatg atcaaatttc gatagtcagt aatactagaa 1560
gaatcatatg tctgaaaata aaagccagaa ccttttctgc ttttgttttc ttttgcccca 1620
agatagggtt tctctcagtg tatccctggc atccctgcct ggaacttctt ttgtaggttt 1680
ggtagcctca aactcagaga ggtcctctct gcctgcctgc ctgcctgcct gcctgcctgc 1740
ctgcctgcct gcctgcctca cttcttctgc caccacaca accgagtcga acctaggatc 1800
tttattttct tctctttctc tcttctttct ttcttctttt ctttctttct ttctttcttt 1860
ctttctttct ttcttattca attagttttc aatgtaagtg tgtgtttgtg ctctatctgc 1920
tgcttatagg cctgcttgcc aggagagggc aacagaacct aggagaaacc accatgcagc 1980
tcctgagaat aagtgaaaaa acaacaaaaa aaggaaattc taatcacata gaatgtagat 2040
atatgccgag gctgtcagag tgctttttta ggcttagtgt aagtaatgaa aattgttgtg 2100
tgtcttttat ccaaacacag aagagaggtg gctcggcctg catgtctgtt gtctgcagtg 2160
agaccaggct ggccttgaac acattaatct gtctgcctct gcttccctaa tgctgcgatt 2220
aaaggcatgt gccaccactg cctctctttt tttcttcttt tttttttttt tggaaaatac 2280
ctttctttct tttctctctc ctcttcttct cttccttctt ttctttctat tctttttttc 2340
tttctttttt cttttttttt ttttttttaa aatttgccct aggttaaagg tgtgctccac 2400
aattgcctca gctctgctct aattctcttt aaaaaaaaac aaacaaaaaa aaaacaaaaa 2460
cagtagtat gtatgtatat ttagaagaaa ttaataactc ttttttctta 2520
aaattcatgt cattcttgtt ccacaaagtg agttccagga cttaccagag aaacctgtgt 2580

-5-

ttcaaatttc	tgtgttcaag	gtcacccctgg	cttacaaagt	gagttccaag	tccgataggg	2640
ctacacagaa	aaaccatata	tcagaaaaaa	aaaaagtcc	aaacacacac	acacacacac	2700
acacacacac	acacacacac	acacacacac	acacacacag	cgccgcgcgc	cgatgagggg	2760
aagtcgtgcc	taaaataaat	atTTTTctgg	ccaaagtga	agcaaatcac	tatgaagagg	2820
tactcctaga	aaaaataaat	acaaacgggc	TTTTtaata	ttccagcaact	gttttaattt	2880
aactctgaat	ttagtcttgg	aaaagggggc	gggtgtgggt	gagtgagggc	gagcgagcag	2940
acgggcgggc	gggcggtgga	gtggccggcg	gcggtggcag	cgagcaccag	aaaacaacaa	3000
accccaagcg	gtagagtgtt	ttaaaaatga	gacctaaatg	tggtggaacg	gaggtcgccg	3060
ccaccctcct	cttccactgc	ttagatgctc	ccttccccct	actgtgctcc	cttccccctaa	3120
ctgtgcctaa	ctgtgcctgt	tccttcaccc	cgctgattcg	ccagcgacgt	actttgactt	3180
caagaacgat	tttgccctgtt	ttcacccgtc	cctgtcatac	tttcgttttt	gggtgcccgga	3240
gtctagcccg	ttcgctatgt	tcgggcgggga	cgatggggac	cgttttgtgoc	actcggggaga	3300
agtgggtgggt	gggtacgctg	ctccgtcgtg	cggtgcgtgag	tgccggaacc	tgagctcggg	3360
agaccctccg	gagagacaga	atgagtgaat	gaatgtggcg	gcgcgtgacg	gatctgtatt	3420
ggtttgtatg	gttgatcgag	accattgtcg	ggcgacacct	agtgtgaca	agtttcggga	3480
acgctccagg	cctctcagggt	tggtgacaca	ggagagggaa	gtgctgtgg	tgaggcgacc	3540
aggggtgacag	gagggcgggc	aagcaggcgg	gagcgtctcg	gagatgggtg	cggtgttaag	3600
gacggctctc	aacgaaggag	tcgtacaggg	agatggccaa	agcagaccga	gttgcgtgtac	3660
gcccttttgg	gaaaaatgct	aggggttggtg	gcaacggtac	taggtcgacc	agaaggctta	3720
agtcctaccc	ccccccccc	tttttttttt	tttctccag	aagccctctc	ttgtccccgt	3780
caccggggggc	acogtacatc	tgaggccgag	aggacgcgat	gggcccgggt	tcacaggcgg	3840
tgtggctcgg	ccagctggcg	cttcgggtct	tttttttttt	tttttttttt	ttttcctcca	3900
gaagccttgt	ctgtcgctgt	caccgggggc	gctgtacttc	tgaggccgag	aggacgcgat	3960
gggccccggc	ttccaaagcg	gtgtggctcg	gccagctgga	gcttcgggtc	tttttttttt	4020
tttttttttt	ttttttttct	cagaagcctt	gtctgtcgct	gtcacggggg	gcgctgtact	4080
tctgaggccg	agaggacgcg	atgggtcggc	ttccaagccg	atgtggcggg	gccagctgga	4140
gcttcgggtt	tttttttttt	ctccagaagc	cctctcttgt	ccccgtcacc	gggggcgctg	4200
tactcttgag	gccgagagga	cgtgatgggc	ccgggttcca	ggcggatgtc	gcccggctag	4260
ctggagcttt	ggatcttttt	tttttttttt	cctccagaag	ccctctcttg	ccccgtcac	4320
cgggggcacc	ttacatctga	gggcgagagg	acgtgatggg	tcgggtctcc	aagccgatgt	4380
ggcggggcca	gctggagctt	cgggtttttt	ttttttctc	cagaagccct	ctctgttccc	4440
cgtcacccgg	ggcgtgttac	ttctgaggcc	gagaggacgt	gatggggccg	ggttccaggc	4500
ggatgtcgcc	cggtcagctg	gagcttttga	tcattttttt	ttttccctcc	agaagccctc	4560
tcttgtcccc	gtcacccggg	gcaccgtaca	tctgaggccg	agaggacacg	atgggcctgt	4620
cttccaagcc	gatgtggccc	ggccagctgg	agcttcgggt	cttttttttt	ttttttctc	4680
cagaagcctt	gtctgtcgct	gtcacccggg	gcgctgtact	tctgaggccg	agaggacgcg	4740
atgggcccgg	cttccaagcc	gggtgtggct	ggccagctgg	agcttcgggt	cttttttttt	4800
tttttttttt	ttctccaga	aaccttgtct	gtcgtgtca	cccggggcgc	ttgtacttct	4860
gatgccgaga	ggacgcgatg	ggcccgctct	ccaggccgat	gtggcccggt	cagctggagc	4920
tttggatctt	tttttttttt	ttttcctcca	gaagccctct	cttgtccccg	tcaccggggg	4980
caccttacat	ctgaggccta	gaggacacga	tgggcccggg	ttccaggccg	atgtggcccg	5040
gtcagtgaga	gttttggatc	tttttttttt	tttttttcca	gaagccctct	tgtccccgtc	5100
accggtggca	ctgtacatct	gaggcggaga	ggacattatg	ggcccggctt	ccaatccgat	5160
gtggcccggt	cagctggagc	tttggatctt	attttttttt	taattttttt	ttccagaagc	5220
cctcttgtcc	ctgtcacccg	tggcacggta	catctgagge	cgagaggaca	ttatggggcc	5280
ggcttccagg	ccgatgtggc	ccggtcagct	ggagctttgg	atcttttttt	ttttttttct	5340
tttttctctc	agaagccctc	tctgtccctg	tcaccggggg	ccctgtacgt	ctgaggccga	5400
gggaaagcta	tgggcgcggg	tttctttcat	tgacctgtcg	gtcttatcag	ttctccgggt	5460
tgtcagggtc	gaccagttgt	tcctttgagg	tcgggttctt	ttcgttatgg	ggtcattttt	5520
gggccacctc	cccaggatag	acttcagggc	gtcgttgctc	gcctgtcact	ttcctccctg	5580
tctcttttat	gcttgtgatc	ttttctatct	gttctctatt	gacctggaga	taggtactga	5640
cacgctgtcc	tttccctatt	aacactaaag	gacactataa	agagaccctt	tcgattttaag	5700
gotgttttgc	ttgtccagcc	tattcttttt	actggcttgg	gtctgtcgcg	gtgctgaag	5760
ctgtccccga	gccacgcttc	ctgctttccc	gggcttgcgt	cttgcgtgtg	cttgcgtgtg	5820
gcagcttgtg	acaactgggc	gctgtgactt	tgtgcgtgtg	cagacgtttt	tcccgatctc	5880
cccaggtgtg	cgttgtcaca	cctgtcccgg	ttggaatggg	ggagccagct	gtggttgagg	5940
gccaccttat	ttcggtcac	tttttttttt	tttttttctc	ttggagtccc	gaacctccgc	6000
tcttttctct	ttccgggtctt	tcttccacat	gcctcccgag	tgcatttctt	tttgtttttt	6060
ttcttttttt	tttttttttt	ttggggaggt	ggagagtcce	gagtacttca	ctcctgtctg	6120
tggtgtccaa	gtgttcatgc	cacgtgcctc	ccgagtgcac	ttttttttgt	ggcagtcgct	6180
cgtttgtgtc	tcttgttctg	tgtctgcccg	tatcagtaac	tgtcttgccc	cgctgtgaag	6240
acatttctat	ctcgttgtgt	tctcccgatt	cgcgctcggt	gctcactctt	agatcgatgt	6300
gggtgtccgg	agttctcttc	gggcccaggc	caagccgcgc	caggcagggg	acggacattc	6360
atggcgaatg	gcggtccgctc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccatc	6420
cgtctgcggg	tggtgtgtgg	aaggcagggg	tgcgggtctc	cgcccccagc	ctgccccgcg	6480
cgcacttttc	tcagtggttc	gcgtgtcct	tgtgtgtgtg	tgaggcgccc	gggtgtgccc	6540
tcacgtgttt	cactttgggtc	gtgtctcgct	tgaccatgtt	cccagagtcg	gtggatgtgg	6600

-6-

ccggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgctgtttct	tgtaagcgtc	6660
gaggtgtctc	tggagcggtc	caggtttgtc	tcctaggtgc	ctgcttctga	gctgggtggtg	6720
gcgctcccca	ttccctgggtg	tgccctccgt	gctccgctctg	gctgtgtgccc	ttcccggtttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtaggggt	ggtgctgagc	6900
ggtcgcggtc	gggggttgaa	agtttctcga	gagactcatt	gctttcccg	ggggagcctt	6960
gagaggcctg	gctttcgggg	gggaccggtt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggt	cgcccgcgtg	7080
ttggctctct	ggtttccctg	tgtgctcgtc	gcctgcctcc	tctctcgggtg	gcccgggctc	7140
gtcgggggtt	tgggtccgtc	cgcctccag	tgagaaagt	tccttctcta	gctatcttcc	7200
ggaaagggtg	cgggcttctt	acggtctcga	gggggtctctc	ccgaatggtc	ccctggagg	7260
ctcgccccct	gaccgcctcc	cgcgcgcgca	gctgttgcctc	tctcgtctac	cgcgcccg	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgcgtc	ggagcatgtg	gctcggcttg	tgtggttggt	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttccctcc	cctcctgagg	gcccgcgtgc	ggacgggggtg	7500
tgggtaggcg	acgggtgggt	cccgggtccc	cacccgtctt	cccgtgcctc	acccgtgcct	7560
tcgctgcgtc	cgtccctct	cgtcgcgtc	cacgactttg	gcccgtcccg	cgacggcgcc	7620
ctgcgcgcgc	cgtggtgcgt	gctgtgtgct	tctcgggtctg	tgtggttgtg	tcgcctcgcc	7680
ccccctctcc	cgcggcgagc	ttcccacggc	tggcgaaatc	gcgggagtc	tccttccct	7740
cctcgggggt	gagagggtcc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacgggacc	7800
gttctgtggg	agaacgggtg	ttggccgctg	cccgcgcgac	gtcggacgtg	gggacccact	7860
gcccgtcggg	ggtcttcgtc	ggtaggcatc	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cggtgtcgcc	tcctcgggt	cccggggggc	cgtcgtgttt	cgggtcggct	cggcgctgca	7980
gggtgtggtg	gactgctcag	gggagtgtg	cagtgtgatt	cccgcgggtt	ttgcctcgcg	8040
tgccctgacc	ggtccgacgc	ccgagcggtc	tctcgggtccc	ttgtgaggac	ccccttccgg	8100
gaggggcccc	tttcggccgc	ccttgccgtc	gtcgcggggc	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgcgcga	gcccgtcttt	ttcctctct	ccccctctct	cctctgaetg	8220
accctgggg	gtgctgtcgg	acccccgca	tgggggcccgc	cgggcacgta	cgctccggg	8280
cggtcaccgg	ggtcttgggg	gggggcccag	gggtaagaaa	gtcggctcgg	cgggcccggag	8340
gagctgtggt	ttggaggggc	tcccggcccc	gcccgcgtgg	cggtgtcttg	cgcggctctg	8400
gagagggtg	cgtgcgagg	gaaaagggtg	cccgcgagg	gcaaagggaa	agaggctagc	8460
agtgggtcatt	gtcccgcagg	tgtgggtggtc	tgttggccga	ggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgcgt	ggttggggcct	gcccggagtc	cgagggtgggt	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgtg	tggcgggtgg	ggctccggtc	8640
gatgtctacc	ttcctctccc	cgaggctctc	ggccttctcc	gcccgggtctc	tcggccctcc	8700
cctcgttctc	cctctcgcg	gggttcaagt	cgtcgtcga	cctccccctc	tcggtccttc	8760
catctctcgc	gcaatggcgc	cgcgcgagtt	cacgggtgggt	tcgtcctcgg	cctccgcttc	8820
tcgcgggggg	ctggcgcgtg	tccggtctct	cctgcgcgac	cccgttggc	gtggtcttct	8880
ctcgcgggtc	tcgcggactc	ctggcttcgc	cggagggtc	agggggcttc	ccggttcccc	8940
gacgttgccg	ctcgtgctg	tgtgcttggg	ggggggcccgc	tgcggcctcc	gcccgcgggt	9000
gagccccctg	cgcaccgcgc	ggtgtgcgtg	ttcgcctcgc	ggtcagttgg	gcccgtcggt	9060
tgtgtcgcgt	cgggagcgtg	ttcgcctcgc	ggcggctaga	cgcgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccagg	ctcgcgcccg	ccgacccccg	cctgcccgtc	ccggtggtgg	9180
tcgttgggtg	ggggagtga	tgggtgtacc	ggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcgggtgt	cgcgcttctc	tttcgcgcaa	ccccacgcgc	aacccaccac	cctgctctcc	9300
cggcccggtg	cggtcgacgt	tccggctctc	ccgatgccga	ggggttcggg	atttgtgcgc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
ggtggctttg	cccggtgcgt	gtgctcgcgg	acgggttttg	tcggaccccc	acgggggtcg	9480
tcggcgca	tgcactctcc	cgttccgcgc	gagcgcggc	ccggctcacc	cccggtttgt	9540
cctccgcgca	ggctctccgc	cgcgcgcgc	tcctcctcct	ctctcgcgt	ctctgtccc	9600
cctggtcctg	tcccaccccc	gacgctccgc	tcgcgttctc	ttacctggtt	gatcctgcca	9660
ggtagcata	gcttgtctca	aagattaaag	catgcatgtc	taagtaacga	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatgggtcc	tttgggtcgt	cgctcctctc	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgacggggcg	tgacccccct	9840
tcccgggggg	ggatgcgtgc	atttatcaga	tcaaaaccaa	cccgttgagc	tccctcccg	9900
ctccggccgg	gggtcggggc	ccggcggtt	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccg	tggcggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcggtgcc	taccatggtg	accacgggtg	acggggaatc	agggttcgat	tccggagagg	10080
gagcctgaga	aacgggtacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
cgacccgggg	aggtagtgc	gaaaaataac	aatacaggac	tctttcgagg	ccctgtaatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctggtgccag	10260
cagccgcgg	aattccagct	ccaatagcgt	cgccgcgagg	tgtcgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcccggcggtc	cgtcgcgagg	cgagtcaccg	cccgtccccg	10380
ccccttgctc	ctcggcgccc	cctcgtatgt	cttagctgag	tgtcccgcgg	ggcccggaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggccgca	gccccttga	taccgagct	10500
aggaataatg	gaataggacc	gcggtcttat	tgtgtggtt	ttcggaactg	aggccatgat	10560
taagagggac	ggccggggggc	attcgtattg	cgcgcgtaga	ggtgaaattc	ttggaccggc	10620

-7-

gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggaggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcggcg	gcgttatctc	catgaccgcg	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggt	tgcaaaagctg	aaacttaaag	gaattgacgg	aagggcacca	10860
ccaggagtg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccggaca	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtggtgg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctgctatg	11040
ctaactagtt	acgcgacccc	cgagcggctg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcg	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgacgcggcg	ctacactgac	tggctcagcg	tgtgcctacc	ctgcgccggc	aggcgcggtg	11220
aaccggttga	accccatctg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtccttccc	ctttgtacac	11340
accgcccgtc	gtactaccgc	attggatggt	ttagtgaggg	cctcggatcg	gccccgcggg	11400
ggtcggccca	cggccctggc	ggagcgtgga	gaagcggctc	gaacttgact	atctagaggga	11460
agtaaaagtc	gtaacaaggt	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcgccgt	ggcccgctct	ccccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgcgtgcgtc	ccgggtcccg	tcgcccgcgt	gtggtagcgag	gtgtctggag	tgaggtgaga	11640
gaagcgcgcg	gtggggtcgg	tctgggtccg	tctgggaccg	cctccgattt	ccccctcccc	11700
tcctctctcc	ctcgtccggc	tctgacctcg	ccacctacc	gcggcgccgg	ctgctcggcg	11760
gcgtcttgcc	tctttcccg	ccggtctctc	cgtgtctacg	agggcgcggt	cgtcgttacg	11820
gggtttttgac	ccgtcccggg	ggcgttcggg	cgtccggcg	cgcgctttgc	tctcccgcca	11880
cccatccccc	ccgcggctct	ggcttttcta	cgttggtcgg	ggcggttgct	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcggccc	tcccgatgcc	acgcttttct	ggcctcgcgt	12000
gtctctcccc	ctcctgtccc	gggtacctag	ctctcgcgtt	ccggcgcgga	ggtttaaggga	12060
ccccgggggg	gtcgccctgc	cgcccccagg	gtcggggggc	gggtggggccc	gtaggggaagt	12120
cgggtcgttcg	ggcggtctct	cctcagactc	catgacctc	ctccccccgc	tgcccgccgtt	12180
cccagggcgg	cggctcgtgtg	gggggggtgga	tgtctggagc	ccccctgggg	gccgtggggg	12240
cccgaccccg	gcgcgcggct	tgcccgatct	ccgcgggtcg	gtcctgtcgg	tgccggtcgt	12300
gggttcccg	gtcgttcccc	tgtttttccg	ctcccgacce	tttttttttc	ctccccccca	12360
cacgtgtctc	gtttcgttcc	tgtgtggccg	cctgaggcta	ccccctcggt	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tcgttggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	ctcacaccgc	aaataccgat	acgactctta	gcggtggatc	actcggtcgt	12540
tgcgctcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttctctc	cggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcaccgcg	tcgggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	acccggtcgt	ggcctccgt	ctcccgaaat	tcagacgtgt	12780
gggcgggttc	cgggtgtggcg	cgcgcgcccc	cgtcgcggag	cctggtctcc	cccgcgcac	12840
cgcgtctcgt	gcttcttccc	gctccgcgt	tcccgccctc	gcccgtgcac	cccgtcctg	12900
gcctcgcgtc	ggcgccctcc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgccccg	12960
tgggaaccca	ccgcgccccc	gtggcgcccc	ggggtggggc	cgtccgcac	tgctctggtc	13020
gaggttggcg	gttgaggggtg	tgctgcgcgc	gaggtggtgg	tcggtcccct	gcggcccggt	13080
ggttgtcggg	gtggcggtcg	acgagggccg	gtcggtcgcc	tgccggtggt	gtctgtgag	13140
gtttgggtct	tgcgctgggg	gaggcggggg	cgaccgctcg	cggggttggt	gcggtcggcc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatccccgc	tgccgttgcg	agggaggggt	tggcgctccg	cgtccgtccg	tccctccctc	13320
cctcgggtggg	cgccttcgcg	ccgcacgcgg	ccgctagggg	cggctcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctcctt	cacccgggcg	gtaccgcgtc	cggcgccggc	13440
ccgcgggacg	ccgcggcgct	cgtgcgcga	tgcgagtcac	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gaccggttgc	gtcccggtt	ccctgggggg	gaccggcggt	13560
ctgtgggctg	tgcgctcccg	gggttgcggt	tgagtaagat	cctccacccc	cgcgcgccct	13620
ccctcccgcc	ggcctctcgg	ggacccctct	agacggttcg	ccggctcgtc	ctcccggtgc	13680
gcccgggtg	gtctctttcc	cgcgcgcctc	ctcgctctct	tcttcccgcg	gctgggcggc	13740
tgtccccctt	ttctgaccgc	gacctcagat	cagacgtggc	gaccgctga	atttaagcat	13800
attagtcagc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcggaatc	cccgcgcgcg	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gacccactcc	cggcgccgcg	tcgtgggggg	cccaagtcct	tctgatcgag	gcccagcccc	13980
tggacgggtg	gaggccggta	gcggccccgg	cgcgcggggc	togggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaactcca	tctaaggcta	aataccggga	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagtgaag	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaaaccgt	taagaggtaa	acgggtgggg	tccgcgciagt	ccgcccggag	14220
gattcaaccc	ggcgccgcgc	gtccggccgt	gcccgggtgt	cccggcggt	ctttcccgct	14280
ccccgttctt	ccgacccctt	ccaccgcgcg	ggtgttcccc	tcttccctcc	cgcgtccggc	14340
gcccctcggc	gcggggcgcg	gggggtggtg	gtgtggtggc	cgcggggcg	gcccgggggtg	14400
gggtcggcgg	gggaccgcgc	ccggccggcg	accggccgcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tccgggacgg	ccgggaaggc	ccggtgggga	aggtggctcg	14520
ggggggggcg	cgcgctcag	ggcgcccgga	accacctcac	cccagtggtt	acagccctcc	14580
ggccgcgctt	tcgccgaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgctctcc	14640

-8-

ctctcccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggccg	ccccccccac	14700
ggcgcgaccg	ctctcccacc	cccccccgtc	gcctctctcg	gggcccgggtg	ggggggcgggg	14760
cggactgtcc	ccagtgcgcc	ccgggctgcg	tccgcgcgtc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcagaa	gccgagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgacccgtct	tgaaacacgg	accaaggagt	ctaaccgcgtg	cgcgagtcag	gggctcgtcc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggccccgcgc	ggggggcccga	ggtgggatcc	15000
cgaggccctc	ccagtccgcc	gagggcgcac	caccggcccg	tctcgcccg	cgcgcgggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	atagggcgga	aagactaatc	gaacatcta	gtagctggtt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgtc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaacctat	tctcaaaactt	taaatgggta	15360
agaagcccgg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgcctag	tgggccaactt	15420
ttggtaagca	gaactggcgc	tgcgggatga	accgaaacgc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agaccccaga	aaaggtgttg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtcggaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gtggagcgt	cgggcccata	cccgcccgtc	gccgcagtcg	gaacggaaacg	15660
ggagcgggagc	ggcgcggggt	gcgcgctctc	cggggtcggg	ggtgcgtggc	ggggggcccg	15720
ccccccgctc	ccctccgcgc	gcccgggttcg	cccccgccgc	gtcgggcccc	gcccggagccta	15780
cgccgcgacg	agtaggagg	ccgctgcggt	gagcctgaa	gcctagggcg	cgggcccggg	15840
tgggagagggc	gcaggtgcag	atcttgggtg	tagtagcaaa	tattcaaacg	agaactttga	15900
aggccgaagt	ggagaagggt	tccatgtgaa	cagcagttga	acatgggtca	gtcggtcctg	15960
agagatgggc	gagtgcggtt	ccgaagggac	gggcgatggc	ctccgttgcc	ctcggccgat	16020
cgaaaggggag	tcggtttcag	atccccgaat	ccggagtggc	ggagatgggc	gcccgcaggc	16080
cagtgcggta	acgcgaccga	tcccgagaa	gccggcgagg	ggcctcgggg	agagtctctc	16140
tttcttttgt	aagggcagg	cgccctggaa	tgggttcgcc	ccgagagagg	ggcccgtgcc	16200
ttggaaagcg	tgcggttcc	ggcgcgctcc	ggtgagctct	cgctggccct	tgaataatccg	16260
ggggagaggg	tgtaaatctc	gcgccggggc	gtaccatata	ccgcagcagg	tctccaaggt	16320
gaacagcctc	tggcatgttg	gaacaatgta	ggtaagggaa	gtcggcaagc	cggatccgta	16380
acttcgggat	aaggattggc	tctaagggtc	gggtcgggtc	ggctggggcg	cgaagcgggg	16440
ctgggcgcg	ccgcggctg	gacgagggc	cgcgcgccctc	tcccacgtcc	ggggagaccc	16500
cccgctcctt	ccgcccgggc	ccgcccctcc	ctcttcccgc	cgggggccccg	tcgctccccg	16560
cgctcgctgc	acctctcttc	ccccctcctt	cttcccgtcg	gggggcccgg	cggggggtcgg	16620
cgcgcgggcg	gggctccggg	gcgccgggtc	caaccccgcg	ggggttccgg	agcgggagga	16680
accagccggtc	cggggtggg	cgggggggcc	ggacactcgg	ggggccggcg	gcccggcgga	16740
ctctggacgc	gagccggggc	cttcccgtgg	atcgctcag	ctgcggcggg	cgctcgcgcc	16800
gctcccgggg	agcccggcg	gtgcccggcg	gggtcccctc	cccgcggggc	ctcgctccac	16860
cccccatcg	cctctccga	ggtgcgtggc	gggggcgggc	gggcgtgtcc	cgcgcgtgtg	16920
gggggaacct	ccgcgtcgg	gttccccgc	cgggtccgcc	ccccggggcg	cggttttccg	16980
cgcgcgcgcc	ccgcctcggc	cggcgccctag	cagccgaact	agaactggtg	cggaccaggg	17040
gaatccgact	gtttaattaa	aacaaagcat	cgcgaaggcc	cgcggcggg	gttgacgca	17100
tgtgatttct	gcccagtgtc	tgtaattgta	aagtgaagaa	attcaatgaa	gcccgggtaa	17160
acggcgggag	taactatgac	tctcttaagg	tagccaaatg	cctcgctatc	taattagtga	17220
cgcgcgatgaa	tggatgaacg	agattcccac	tgtccctacc	tactatccag	cgaaccaca	17280
gccaagggaa	cgggcttggc	ggaaatcagc	gggaaagaa	accctgttga	gcttgactct	17340
agtctggcac	ggtgaagaga	catgagaggt	gtagaataag	tgggaggccc	ccggcgcccc	17400
gccccgtcct	cgcgtcgggg	tgggggcacg	ccggcctcgc	gggcccggcg	tgaataacca	17460
ctactctcat	cgttttttca	ctgacccgg	gaggcggggg	ggcgagcccc	gaggggctct	17520
cgcttctggc	gccaagcgtc	cgtcccgcgc	gtgcccggcg	gcccgaaccc	ctccggggac	17580
agtgccagg	ggggagtgtg	actggggcgg	tacacctgtc	aaacggtaac	gcaggtgtcc	17640
taaggcgagc	tcagggagga	cagaaacctc	ccgtggagca	gaagggcaaa	agctcgctct	17700
atcttgattt	tcagtagggt	tacagaccgt	gaaagcgggg	cctcacgata	cttctgacct	17760
tttgggtttt	aagcaggagg	tgtcagaaaa	gttaccacag	ggataactgg	cttgtggcg	17820
ccaagcggtc	atagcgacgt	cgctttttga	tctctcgatg	tccgctcttc	ctatcattgt	17880
gaagcagaat	tcaccaagcg	ttggattgtt	cacctactaa	tagggaacgt	gagctgggtt	17940
tagaccgtcg	tgagacaggt	tagttttacc	ctactgatga	tgtgtgtgtg	ccatggtaat	18000
cctgctcagt	acgagaggaa	ccgcaggttc	agacatttgg	tgtatgtgct	tggctgagga	18060
gccaatgggg	cgaagctacc	atctgtggga	ttatgactga	acgcctctaa	gtcagaatcc	18120
gccaagcgg	aacgatacgg	cagcgccgaa	ggagcctcgg	ttggcccggg	atagccgggt	18180
ccccgtccgt	cccgtcggc	ggggtccccg	cgtcgccccg	cggcgccgcg	gggtctcccc	18240
ccgcccggcg	tgggacccgg	ggtccgggtc	ggagagccgt	tccgtctggg	aaacgggggtg	18300
cggccggaaa	gggggcccgc	ctctgccccg	tcacgttgaa	cgcacgttcg	tgtggaacct	18360
ggcgctaaac	cattcgtaga	cgacctgctt	ctgggtcggg	gtttcgtagc	tagcagagca	18420
gctccctcgc	tgcgatctat	tgaagtcag	ccctcgacac	aagggtttgt	ctctcggggc	18480
tttcccgctc	cagccccgct	cgctcgcacg	cgaccgtgtc	gcccgggggg	cgtcacgggg	18540
gcggtcgctc	cggccccgc	gagggacggt	gtgggtgggtg	gtgggtgggtg	ggggggggat	18600
cgtcttctcc	tccgtctccc	gagggacggt	cgtttctctt	cccccttccg	tccgtctctc	18660

-9-

tgggtgtggg	agcctcgtgc	cgtcgcgacc	gcggcctgcc	gtcgcctgcc	gccgcagccc	18720
cttgccctcc	ggccttggcc	aagccggagg	gcggaggagg	gggatcggcg	gcggcggcga	18780
ccgcggcgcg	gtgacgcacg	gtgggatccc	catcctcggc	gcgtccgtcg	gggacggccg	18840
gttgaggagg	cgaggagggt	ttttcccggt	aacgcgcggt	tcggcgccag	gcctctggcg	18900
gcgggggggg	cgctctctcc	gcccagagcat	ccccactccc	gcccctcctc	ttcgcgcgccc	18960
gcggcgggcga	cgtgcgtacg	aggggaggat	gtcgcgggtgt	ggaggcggag	aggggtccggc	19020
gcggcgccctc	ttccattttt	tcccccccaa	cttcggagggt	cgaccagtac	tccggcgac	19080
actttgtttt	ttttttttcc	cccgatgctg	gaggtcgacc	agatgtccga	aagtgtcccc	19140
cccccccccc	ccccccggcg	cggagcggcg	gggccaactct	ggactctttt	tttttttttt	19200
tttttttttt	ttaaattcct	ggaaccttta	ggtcgaccag	ttgtccgtct	tttactcctt	19260
catatagggtc	gaccagtact	ccgggtggta	ctttgtcttt	ttctgaaaat	cccagagggtc	19320
gaccagatat	ccgaaagtcc	tctctttccc	tttactcttc	cccacagcga	ttctcttttt	19380
tttttttttt	tttgggtgtgc	ctctttttga	cttatataca	tgtaaatagt	gtgtacgttt	19440
atatacttat	aggaggaggt	cgaccagtac	tccgggcgac	actttgtttt	tttttttttt	19500
tccaccgatg	atggagggtcg	accagatgtc	cgaaagtgtc	ccgtccccc	cctccccccc	19560
ccgcgacgcg	gcggggtcac	tctggaactct	tttttttttt	tttttttttt	tttaaatctc	19620
tggaaacctta	aggtcgacca	gttgtccgtc	tttcaactcat	tcataatagg	cgaccgggtgg	19680
tactttgtct	ttttctgaaa	atcgcagagg	tcgaccagat	gtcagaaagt	ctggtgggtcg	19740
ataaattatc	tgatctagat	ttgtttttct	gtttttcagt	tttgtgtgtg	tttgtgtgtg	19800
tttgtgtgtg	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	19860
tttgtgtgtg	gttgtgtgtg	gttgtgtgtg	gttgtgtgtg	gttgtgtgtg	gttgtgtgtg	19920
gttgtgtgtg	gttgtgtgtg	gttgtgtgtg	gttgtgtgtg	gttgtgtgtg	gttgtgtgtg	19980
ttgtttgtctg	ttgtttgtgtg	tttgtgggtg	tttgtgtgtg	tttgtgtgtg	tttgtgtgtg	20040
tacacaaaca	tgcacttttt	ttaaataaaa	tttgttttgt	tttgttttgt	tttgttttgt	20100
tatcccttct	cttctctctc	ttttttaaaa	attttctttg	tgtgtgtgtg	tgtgtgtgtg	20160
tgtgtgtgtg	tgcgtgtgtg	tgtgtgtgtg	cgtgcagcgt	gcgcgcgctc	gttttataaa	20220
tacttataat	aataagtcgc	cgggtgggtg	tagcttcccc	gactccagag	gcagaggcag	20280
gcagacttct	aggttcgagg	ccagcctggg	ctacagaggg	accctgtctc	gaaaaatgaa	20340
aataaataca	tacatacata	catatacata	tacatacata	catatacata	tacatacata	20400
gttgaccagt	tgtcaatcct	ttagaatttt	gtttttaatt	aatgtgatag	agagatagat	20460
aatagataga	tggtatagat	gatacaaaata	taggtttttt	tttcagttaa	tatgagggtg	20520
attaaccact	tttccctttt	taggtttttt	tttttttccc	ctgtccatgt	ggttgctggg	20580
atttgaactc	aggaccctgg	cagggtcaact	ggaaaacgtg	ttttctatat	atataaatag	20640
tgggtctgtct	gctgtttgtt	tgtttgtctt	cttctgtgta	ctgtctgtct	ctgtctgtct	20700
tgcttttttt	ttctcttga	caatcgaact	tgcttggtct	gatttggtct	gatttggtct	20760
tctgtagacc	agcctggcct	gctacaccac	agttgggtct	gatttggtct	gatttggtct	20820
caatttttga	gtaaagggtg	gaaatcaact	ggaggtttct	tttgtttccg	atttggtgtg	20880
attttattat	tagacagaac	ggctgggttt	taactcgaac	gaatgcagta	agggttaagg	20940
attagttgga	ccaattagtt	ttctctagct	ttctctctct	gaggaaaaaa	taaaaataata	21000
ttgtgggggt	ggggatcagg	ttctctagct	ttctctctct	ttctctctct	ttctctctct	21060
gattttttgt	aagattactt	ttctctctct	ttctctctct	ttctctctct	ttctctctct	21120
ttcattgtct	ctgccttctg	gaagatgtag	agttttggat	tcaagccgta	taattgtttat	21180
ccagttcctc	tgaaccagag	agttttggat	tttttttttt	tttttttttt	tttttttttt	21240
gatgtgctga	tgaaccagag	agttttggat	tttttttttt	tttttttttt	tttttttttt	21300
aaaagttcta	acaaagtgat	cgcgctttgt	gggttttgctt	aataaaaatgg	tgagctttgt	21360
tggtctcact	ctgccaccaa	gggttttgctt	tggtgtgttt	tggtgtgttt	tggtgtgttt	21420
agacatatat	tttttctttt	gggttttgctt	tggtgtgttt	tggtgtgttt	tggtgtgttt	21480
ttcccagacg	gccttttgag	tggtgtgttt	tggtgtgttt	tggtgtgttt	tggtgtgttt	21540
caccacaact	ctaaccgttt	tggtgtgttt	tggtgtgttt	tggtgtgttt	tggtgtgttt	21600
ggaagagcat	gtagcagttg	taggacacac	gtggttgctt	gagccatctc	agacagtctg	21660
ggttgtgaac	caccaccat	gtggttgctt	gagccatctc	agacagtctg	aagtctccac	21720
agtcagggtc	ctaaaccgat	agacagtctg	aagtctccac	cccacactgc	gaggggtggg	21780
tgaatgatcc	cagcatggga	aagtctccac	cccacactgc	gaggggtggg	gtgggggcag	21840
ataaaaataat	gaaatgaatg	cccacactgc	gaggggtggg	gtgggggcag	tctctgaaat	21900
ctcacagcac	ctccccctcc	gaggggtggg	gtgggggcag	tctctgaaat	ccctgagtg	21960
gaggggtggg	gtgggggcag	tctctgaaat	ccctgagtg	gaggggtggg	gtgggggcag	22020
atggcctggg	tctctgaaat	ccctgagtg	gaggggtggg	gtgggggcag	tctctgaaat	22080
tacctgaagt	ccctgagtg	gaggggtggg	gtgggggcag	tctctgaaat	ccctgagtg	22118

<210> 19
 <211> 175
 <212> DNA
 <213> Mus musculus

<400> 19
 ctcccgcgcg gcccgcgtgt tgcgcgttcc cgtggcgcg acaatgcggt tgtgcgtcca
 cgtgtgcgtg tccgtgcagt gccgttgtgg agtgcctcgc tctcctcctc ctccccggca

60
 120

-10-

gcgttcccac ggttggggac caccggtgac ctgcacctct tcgggcctgg atccg 175

<210> 20
 <211> 755
 <212> DNA
 <213> Mus musculus

<400> 20
 ggtctggttg gaattgttga cctcgtcttc ggggtgcggcc tttggggaac ggcgggggtcg 60
 gtcgtgcccg gcgcgggac tgtgtcgggg ccacttccc gctcgagggt ggcggtgggc 120
 gcggcggttg tagtctccc tgttgctgtc tcccgggctc ttgggggggg tgccgtcgtt 180
 ttcggggccg gcgttgcttg gcttacgcag gcttggtttg ggactgcctc aggagtcgtg 240
 ggcggtgtga ttccgcggg ttttgccctc cgtctgcctg ctttgccctc ggtttgcttg 300
 gttcgtgtct cgggagcgtt ggtttttttt tttttcgggt cccggggaga ggggtttttc 360
 cgggggaagt tcccgctgcc ccctgccgcc ggtgggtttt cgtttcgggc tgtgttcgtt 420
 tcccttccc cgtttcggc tcggttctcc ccggtcggtc ggccctctcc ccggtcggtc 480
 gccggccgt gctgccggac cccctctct gggggggatg cccgggcacg caccggtccg 540
 gccggccact gtggtccggg agctgctcgg caggcgggtg agccagttgg aggggcgtca 600
 tgcccccgcg ggctcccggt gccgacgcgg cgtgttcttt gggggggcct gtgcgtgcgg 660
 gaaggctgcg cactgtgtcg gtccctgcga gggaaagagg cttttttttt ttaggggggtc 720
 gtccttcgtc gtcccgtcgg cgggtgatcc ggcct 755

<210> 21
 <211> 463
 <212> DNA
 <213> Mus musculus

<400> 21
 ggcgagggtg cgtctgcggg ttggggctcg tccggcccg tcgtcctccg ggaaggcgtt 60
 tagcgggtac cgtcgcgcgg ccgaggtggg cgcacgtcgg tgagataacc ccgagcgtgt 120
 ttctggttgt tggcggcggg ggctccggtc gatgtcttcc cctccccctc tcccgaggc 180
 caggctcagcc tccgcctgtg ggcttcgtcg gcgctctccc cccctctcac gtccctcggc 240
 agcgagcccg tccgttcgac cttccttccg ccttcccccc atctttccgc gtcgcgttgg 300
 ccccggggtt ttcaaggcgc ccccaacgct cctccgcctc tccgcccgtg gtttggacgc 360
 ctgggttcgg tctccccggc aaaccccggt tgggttggtc tccggcccg gcttgcctct 420
 cgggtctccc aacccccggc cgggaagggt cgggggttcc ggg 463

<210> 22
 <211> 378
 <212> DNA
 <213> Mus musculus

<400> 22
 ggattcttca ggattgaac ccaaaccggt tcagtttctc ttccggctcc ggcggggggg 60
 ggcggccccc ggcgggttgg tgagttagat aacctcgggc cgatcgcaac ccccccgtgg 120
 cggcgacgac ccattcgaac gtctgcctca tcaactttcg atggtagtgc atgtgcctac 180
 catggtgacc acgggtgacg gggaatcagg gttcgattcc ggagagggag cctgagaaac 240
 ggctaccaca tccaaggaa gcagcaggcg cgcaaattac cactcccga cccggggagg 300
 tagtgacgaa aaataacaat acaggactct ttcgaggccc tgaattgga atgagtcac 360
 tttaaatcct ttaagcag 378

<210> 23
 <211> 378
 <212> DNA
 <213> Mus musculus

<400> 23
 gatccattgg agggcaagtc tggtgccagc agccgcggta attccagctc caatagcgtg 60
 tattaagtt gctgcagtta aaaagctcgt agttggatct tgggagcggg cgggcgggtc 120
 gccgcgaggc gactcaccgc ccgtccccgc cccttgccctc tcggcgcccc ctgatgctc 180
 ttagctgagt tgtcccgcgg ggcgcgaagc gtttactttg aaaaaattag agttgtttca 240
 aagcagggcc gatccgctg gataccgcca gctaggaatg aatggaatag gaccgcggtt 300
 cctattttgt ttggttttcg gaactgagcc catgattaag ggaaacggcc gggggcattc 360
 ccttattgcg ccccccta 378

<210> 24
 <211> 719

-11-

<212> DNA

<213> Mus musculus

<400> 24

ggatctttcc	cgctccccgt	tctctccggc	ccctccaccc	gcgcgtctcc	ccccctcttt	60
tccccctctc	ggaggggggg	gaggtggggg	cgcggtggcg	gggtcggggg	tgggggcggc	120
gggggaccgc	ccccggccgg	caaaaggccg	ccgcggggcg	cacttcaacc	gtagcgggtg	180
gccgcgaccg	gctacgagac	ggctgggaag	gcccgcggcg	gaatgtggct	cggggggggg	240
ggcgcgctct	agggcgcgcc	gaaccacctc	accccgagtg	ttacagccct	ccggccgcgc	300
tttcgcggaa	tcccgggggc	gaggggaagc	ccgatacccg	tcgcccgcgt	tttccccctc	360
ccccgtccgc	ctcccgggcg	ggcggtgggg	tgggggcccg	gccgcccctc	ccacgcccgt	420
ggttttctct	tctcccggtc	tcggccgggt	tggggggggg	agcccgggtg	ggggcggggc	480
ggactgtcct	cagtgcgccc	cgggcgctcg	cgcgccgtcg	ggcccggggg	gttctctcgg	540
tcacgcgcgc	cccgacgaag	ccgagcgcac	ggggtcggcg	gcgatgtcgg	ctacccaccc	600
gacccgtctt	gaaacacgga	ccaaggagtc	taacgcgtgc	gcgagtcagg	ggctcgcacg	660
aaagccgccc	tggcgcaatg	aaggtgaagg	gccccgtccg	ggggcccgag	gtgggatcc	719

<210> 25

<211> 685

<212> DNA

<213> Mus musculus

<400> 25

cgaggcctct	ccagtccgcc	gagggcgcac	caccggcccc	tctcgccgcg	cgcgtcgggg	60
aggtggagca	cgagcgtagc	cgtaggacc	cgaaagatgg	tgaactatgc	ctgggacggg	120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcgggtc	tgacgtgcaa	atcggtcgtc	180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	240
tttccctcag	gatagctggc	gctctcgcaa	ccttcggaag	cagttttatc	cgggtaaaag	300
cggaatggat	taggaggtct	tggggccgga	aacgatctca	aactattttc	caaactttta	360
atgggtaagg	aagcccggct	cgctggcggt	gagccggggc	tggaatgcga	gtgcctagtg	420
ggccactttt	ggtaagcaga	actggcgctg	cgggatgaac	cgaacgcccg	gttaaggcgc	480
ccgatgccga	cgctcatcag	acccagaaa	aggtgttggt	tgatatagac	agcaggacgg	540
tggccatgga	agtcggaatc	cgctaaggag	tgtgtaacaa	ctcacctgcc	gaatcaacta	600
gccctgaaaa	tggatggcgc	tggagcgctc	ggcccatacc	cggccgtcgc	cggcagtcgg	660
aacgggacgg	gacgggagcg	gccgc				685

<210> 26

<211> 5162

<212> DNA

<213> Artificial Sequence

<220>

<223> Chimeric bacterial plasmid

<400> 26

gacggatcgg	gagatctccc	gatcccctat	ggtcgactct	cagtacaatc	tgctctgatg	60
ccgcatagtt	aagccagtat	ctgctccctg	cttgtgtgtt	ggaggtcgct	gagtagtgcg	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgaccga	caattgcatg	aagaatctgc	180
ttagggttag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccggc	tggctgaccg	cccaacgacc	360
cccgcccat	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggt	aaactgccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catggtgatg	cggttttggc	agtacatcaa	tggcggtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	ccatttgacg	caaattggcg	780
gtaggcgtgt	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttggtacc	900
gagctcggat	cgatatctgc	ggccgcgtcg	acggaattca	gtggatccac	tagtaacggc	960
cgccagtggt	ctggaattaa	ttcgctgtct	gcgagggcca	gctgttgggg	tgagtactcc	1020
ctctcaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	1080
ttgatattca	cctggcccgc	gggtgatgct	ttgaggggtg	ccgcgtccat	ctgggtcagaa	1140
aagacaatct	tttgttggtc	aagcttgagg	tgtggagcgc	ttgagatctg	gccatacact	1200
tgagtgacaa	tgacatccac	tttgcctttc	tctccacagg	tgtccactcc	caggtccaac	1260
tgcaggtcga	gcatgcatct	agggcgggcca	attccgcccc	tctccctccc	ccccccctaa	1320

-12-

cgttactggc	cgaagccgct	tgggaataagg	ccggtgtgcg	tttgtctata	tgtgattttc	1380
caccatattg	ccgtcttttg	gcaatgtgag	ggcccggaaa	cctggccctg	tcttcttgac	1440
gagcattcct	aggggtcttt	cccctctcgc	caaaggaatg	caaggtctgt	tgaatgtcgt	1500
gaaggaagca	gttcctctgg	aagcttcttg	aagacaaaca	acgtctgtag	cgaccctttg	1560
caggcagcgg	aaccccccac	ctggcgacag	gtgcctctgc	ggccaaaagc	cacgtgtata	1620
agatacacct	gcaaaaggcg	cacaacccca	gtgccacgtt	gtgagttgga	tagttgtgga	1680
aagagtcaaa	tgggtctcct	caagcgtatt	caacaagggg	ctgaaggatg	ccgagaagg	1740
accccatgtt	atgggatctg	atctggggcc	tgggtgcaca	tgctttacat	gtgttttagtc	1800
gagggttaaaa	aaacgtctag	gccccccgaa	ccacggggac	gtggttttcc	tttgaaaaac	1860
acgatgataa	gcttgaccaca	acccgggatc	caccggtcgc	caccatgggtg	agcaagggcg	1920
aggagctggt	caccgggggtg	gtgcccaccc	tggtcgagct	ggacggcgac	gtaaacggcc	1980
acaagttcag	cgtgtccggc	gagggcgagg	gcgatgccac	ctacggcaag	ctgaccctga	2040
agttcatctg	caccaccggc	aaagctgccc	tgccctggcc	caccctcgtg	accaccctga	2100
cctacggcgt	gcagtgtctc	agccgctacc	ccgaccacat	gaagcagcac	gacttcttca	2160
agtcggccat	gcccgaaggc	tacgtccagg	agcgcacccat	cttcttcaag	gacgacggca	2220
actacaagac	cgcgcccgag	gtgaagtctg	agggcgacac	cctggtgaac	cgcatcgagc	2280
tgaagggcat	cgacttcaag	gaggacggca	acatctgggg	gcacaagctg	gagtacaact	2340
acaacagcca	caacgtctat	atcatggccg	acaagcagaa	gaacggcatc	aaggtgaact	2400
tcaagatccg	ccacaacatc	gaggacggca	gcgtgcagct	cgccgaccac	taccagcaga	2460
acacccccat	cggcgacggc	cccggtgctgc	tgcccgacaa	ccactacctg	agcaccaggt	2520
ccgcccctgag	caaaagacccc	aacgagaagc	gcgatcacat	ggctcctgctg	gagttcgtga	2580
ccgcccgcgg	gatcactctc	ggcatggacg	agctgtacaa	gtaaacggcg	cctagagctc	2640
gctgatcagc	ctcgactgtg	cctctagttg	ccagccatct	gttgtttggc	cctcccccg	2700
gcttctcttg	accctggaag	gtgccactcc	cactgtcctt	tcctaataaaa	atgaggaagt	2760
tgcatcgcat	tgtctgagta	ggtgtcattc	tattctgggg	ggtgggggtg	ggcaggacag	2820
caaggggggag	gattgggaag	acaatagcag	gcatgctggg	gatgcgggtg	gctctatggc	2880
ttctgaggcg	gaaagaacca	gctggggctc	gagtgcattc	tagttgtggg	ttgtccaaac	2940
tcatcaatgt	atcttatcat	gtctgtatac	cgtcgacctc	tagctagagc	ttggcgtaat	3000
catggtcata	gctgtttcct	gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	3060
gagccggaag	cataaagtgt	aaagcctggg	gtgcctaatt	agtgcataa	ctcacattaa	3120
ttgcgttgcg	ctcactgcc	gctttccagt	cggaacacct	gtcgtgccag	ctgcattaat	3180
gaatcggcca	acgcgcgggg	agaggcggtt	tgcgatttgg	gcgctcttcc	gcttctcgc	3240
tcaactgactc	gctgcgctcg	gtcgttcggc	tgccggcgagc	ggtatcagct	cactcaaaag	3300
cggtaatacg	ggtatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaa	3360
gccagcaaaa	gcagcaggaa	cgtaaaaaag	ccgcttggct	ggcgtttttc	cataggctcc	3420
gccccctga	cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aaccgcagag	3480
gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	3540
cctgcgcgt	taccggatac	ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	3600
aatgctcacg	ctgtagggtat	ctcagttcgg	tgtaggctcg	tcgctccaag	ctgggctgtg	3660
tgacgaacc	ccccgttcag	cccgaccgct	gcgcttcttc	cggtaaactat	cgtcttgagt	3720
ccaacccggt	aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	3780
gagcgaggta	tgtagcggt	gctacagagt	ttctcgaagt	gtggcctaac	tacggctaca	3840
ctagaaggac	agtatttgg	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	3900
ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	cggtgggttt	tttggtttga	3960
agcagcagat	aaaaaaggat	ctcaagaaga	ctcaagaaga	tcctttgatc	ttttctacgg	4020
ggtctgacgc	tcagtggaa	gaaaactcac	gttaagggat	tttggctcatg	agattatcaa	4080
aaaggatctt	cacctagatc	cttttaaat	aaaaatgaag	ttttaaatca	atctaaagta	4140
tatatgagta	aacttggctc	gacagttacc	aatgcttaat	cagtgaaggca	cctatctcag	4200
cgatctgtct	atctcgttca	tccatagttg	cctgactccc	cgtcgtgtag	ataactacga	4260
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	4320
cggtctccaga	tttatcagca	ataaacccagc	cagccggaag	ggccgagcgc	agaagtggctc	4380
ctgcgaacttt	atccgcctcc	atocagtcta	ttaatgtgtg	ccgggaagct	agagtaagta	4440
gttcgccagt	taatagtttg	cgcaacggtg	ttgccattgc	tacaggcatc	gtgggtgtcac	4500
gctcgtcggt	tggtatggct	tcattcagct	ccggtttcca	acgatcaagg	cgagttacat	4560
gatcccccat	gttgtgcaaa	aaagcggtta	gtccttccgg	tcctccgatac	gttgtcagaa	4620
gtaagttggc	gcagtggtta	tcactcatgg	ttatggcagc	actgcataat	tctcttactg	4680
tcatgccatc	cgtaagatgc	ttttctgtga	ctggtgagta	ctcaaccaag	tcattctgag	4740
aatagtgat	gcggcgaccc	agttgctctt	gccccgcgtc	aatacgggat	aataccgcgc	4800
cacatagcag	aactttaaaa	gtgctcatca	ttggaaaaacg	ttcttcgggg	cgaaaaactct	4860
caaggatctt	accgctgttg	agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	4920
cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	4980
ccgcaaaaaa	gggaataagg	cgcacacgga	aatgttgaat	actcatactc	ttcctttttc	5040
aatattattg	agcattttat	cagggttatt	gtctcatgag	cggatacata	tttgaatgta	5100
tttagaaaaa	taaacaataa	ggggttccgc	gcacatttcc	ccgaaaaagtg	ccacctgacg	5160
tc						5162

-13-

<211> 5627

<212> DNA

<213> Artificial Sequence

<220>

<223> pMG plasmid from InvivoGen; IRES sequence modified
EMCV nucleotides 2736-3308

<400> 27

caccggcgaa	ggaggcctag	atctatcgat	tgtacagcta	gctcgacatg	ataagataca	60
ttgatgagtt	tggacaaacc	acaactagaa	tgcagtga	aaaatgcttt	atgtgtgaaa	120
tttgtgatgc	tattgcttta	tttgtgaaat	ttgtgatgct	attgctttat	ttgtaaccat	180
tataagctgc	aataaacaag	ttaacaacaa	caattgcatt	cattttatgt	ttcagggttca	240
gggggaggtg	tgggaggttt	tttaaagcaa	gtaaaacctc	tacaaatgtg	gtagatccat	300
ttaaatgtta	attaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	360
ggccgcgttg	ctggcgcttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	420
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttcccc	480
tggagctcc	ctcgtgcgct	ctcctgttcc	gacctgccc	cttaccggat	acctgtccgc	540
ctttctccct	tccggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	600
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccggtc	agccccagcg	660
ctgcgcctta	tccggttaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	720
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	780
gttcttgaag	tgggtggccta	actacggcta	cactgaaga	acagtatttg	gtatctcgcc	840
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	900
caccgctggt	agcgtgtggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	960
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtggg	acgaaaactc	1020
acgttaaggg	attttggcca	tggctaagtt	attaagctgc	aataaacaat	cattattttc	1080
attggatctg	tgtgttggtt	ttttgtgtgg	gcttggggga	gggggaggcc	agaatgactc	1140
caagagctac	aggaaggcag	gtcagagacc	ccactggaca	aacagtggct	ggactctgca	1200
ccataacaca	caatcaacag	gggagtgcgc	tggatcgagc	tagagtccgt	tacataaact	1260
acggtaaatg	gcccgcctgg	ctgaccgccc	aacgaccccc	gcccattgac	gtcaataatg	1320
acgtatgttc	ccatagtaac	gccaataggg	actttccatt	gacgtcaatg	gggtggagat	1380
ttacggtaaa	ctgcccactt	ggcagtacat	caagtgtatc	atatgccaa	tacgccccct	1440
attgacgtca	atgacggtaa	atggcccgcc	tggcatttat	cccagtacat	gaccttatgg	1500
gactttccta	cttggcagta	catctacgta	ttagtcatcg	ctattaccat	gggtgatgcg	1560
ttttggcagt	acatcaatgg	gcgtggatag	cggtttgact	cacggggatt	tccaagtctc	1620
caccocattg	acgtcaatgg	gagtttgttt	tggcaccaaa	atcaacggga	ctttccaaaa	1680
tgtcgtaaca	actccgcccc	attgacgcaa	atgggcggtg	ggcgtgtacg	gtgggaggtc	1740
tataaagca	gagctcgttt	agtgaaccgt	cagatcgcc	ggagacgcca	tccacgctgt	1800
tttgacctcc	atagaagaca	ccgggaccga	tccagcctcc	gcgcccgga	acggtgcatt	1860
ggaacgcgga	ttccccgtgc	caagagtgc	gtaagtaccg	cctatagagt	ctataggccc	1920
accccccttg	cttcttatgc	atgctatact	gtttttggct	tggggtctat	acacccccgc	1980
ttcctcatgt	tataggtgat	ggatagcttt	agcctatagg	tgtgggttat	tgaccattat	2040
tgacctacc	cctattggtg	acgatacttt	ccattactaa	tccataacat	ggctctttgc	2100
cacaactctc	tttattggct	atatgccaat	acactgtcct	tccagagactg	acacggactc	2160
tgtattttta	caggatgggg	tctcatttat	tatttacaaa	ttcacatata	caacaccacc	2220
gtccccagtg	cccgcagttt	ttattaaaca	taacgtggga	tctccacgcg	aatctcgggt	2280
acgtgttccg	gacatgggct	cttctccggt	agcggcgagg	cttctacatc	cgagccctgc	2340
tcccatgcct	ccagcgactc	atggctcgct	ggcagctcct	tgtctcctaac	agtggaggcc	2400
agacttaggc	acagcacgat	gcccaccacc	accagtgtgc	cgcacaaggc	cgtggcggtg	2460
gggtatgtgt	ctgaaaatga	gctcggggag	cgggcttgca	ccgctgacgc	atttggaaga	2520
cttaaggcag	cggcagaaga	agatgcaggc	agctgagttg	ttgtgttctg	ataagagtca	2580
gaggtaaact	ccgttgccgt	gctgttaacg	gtggaggggc	gtgtagtctg	agcagtagtc	2640
gttgctgccg	cgcgcgccac	cagacataat	agctgacaga	ctaacagact	gttctcttcc	2700
atgggtcttt	tctgcagtca	cccgggggat	ccttcgaacg	tagctctaga	ttgagtcgac	2760
gttactggcc	gaagccgctt	ggaataaggc	cgggtgtgct	ttgtctatat	gttattttcc	2820
accatattgc	cgtcttttgg	caatgtgagg	gcccggaaac	ctggccctgt	cttcttgacg	2880
agcattccta	ggggctcttc	ccctctcgcc	aaagggaatg	aaggctctgt	gaatgtcggt	2940
aaggaagcag	ttcctctgga	agcttcttga	agacaaacaa	cgtctgtagc	gaccctttgc	3000
aggcagcgga	acccccacc	tggcgacagg	tgcctctgcg	gccaaaagcc	acgtgtataa	3060
gatacacctg	caaaggcgcc	acaaccccag	tgccacgttg	tgaagtggtg	agttgtggaa	3120
agagtcaaat	ggctctctc	aagcgtattc	aacaaggggc	tgaaggatgc	ccagaaggta	3180
ccccattgta	tgggatctga	tctggggcct	cgggtgcacat	gctttacatg	tgttttagtcg	3240
agggttaaaaa	aacgctctagg	ccccccgaac	cacggggagc	tgggttttct	ttgaaaaaca	3300
cgataataacc	atgggttaagt	gatatactact	agttgtgacc	ggcgccatgt	gttgacaatt	3360
aatcatcggc	atagtatatc	ggcatagtat	aatacgactc	actataggag	ggccaccatg	3420
tcgactacta	accttcttct	ctttcctaca	gctgagatca	ccggtaggag	ggccatcatg	3480

-14-

```

aaaaagcctg aactcaccgc gacgtctgtc gccaagtttc tgatcgaaaa gttcgacagc 3540
gtctccgacc tgatgcagct ctccggagggc gaagaatctc gtgctttcag cttcgatgta 3600
ggagggcgctg gataatgtcct gcgggtaaat agctgcgcgc atggtttcta caaagatcgt 3660
tatgtttatc ggcactttgc atcggccgcg ctcccgatcc cggaggtgct tgacattggg 3720
gaattcagcg agagcctgac ctattgcatc tcccgcctg cagagggtgt caggttgcaa 3780
gacctgcctg aaaccgaact gcccgcgtgt ctgcaaccgc tcgaggagct catggatgcg 3840
atcgctgcgg ccgatcttag ccagacgagc ggggttcggcc cattcggaac gcaaggaatc 3900
ggccaataca ctacatggcg tgatttcata tgcgcgattg ctgatcccca tgtgtatcac 3960
tggaactg tgatggagca caccgtcagt gcgtccgtcg cgcaggctct cgtatgagctg 4020
atgctttggg ccgaggactg ccccgaaagtc cggcacctcg tgcacgcgga tttcggtccc 4080
aacaatgtcc tgacggacaa tggccgcata acagcgggtc ttgactggag cggaggcgatg 4140
ttcggggatt cccaatacga ggtcgccaac atcttcttct ggaggccgtg gttggcttgt 4200
attggagcgc agacgcgcta cttcgagcgg aggcattccg agcttgcagg atcgccgcgg 4260
ctccgggcgt atatgctccg cattggtctt gaccaactct atcagagctt ggttgacggc 4320
aatttcgatg atgcagcttg ggcgcagggt cgtatgcgac caatcgtccg atccggagcc 4380
gggactgtcg ggcgtacaca aatcgcccgc agaagcgcgg ccgtctggac cgtatggctgt 4440
gtagaagtac tcgccgatag tggaaaccga cgcgccagca ctgctccgag ggcaaaggaa 4500
tgagtcgaga attcgctaga gggccctatt ctatagtgtc acctaatgc tagagctcgc 4560
tgatcagcct cgactgtgcc ttctagttag cagccatctg ttgtttgccc ctcccccggtg 4620
ccttccttga ccttggaggg tgcactccc actgtccttt cctaataaaa tgaggaaatt 4680
gcacgcatt gtctgagtag gtgtcattct attctggggg gtggggtggg gcaggacagc 4740
aagggggagg atttgggaaga caatagcagg catgcgcagg gcccaattgc tcgagcggcc 4800
gcaataaaaat atctttatatt tcattacatc tgtgtgttgg ttttttgtgt gaatcgtaac 4860
taacatacgc tctccatcaa aacaaaacga acaaaaacaa actagcaaaa taggctgtcc 4920
ccagtgcagg tgcagggtgcc agaacatttc tctatcgaag gatctgcgat cgctccgggtg 4980
cccgtcagtg ggcagagcgc acatcgccca cagtcgccga gaagtggggg ggaggggtcg 5040
gcaattgaac cgtgtcctag agaaggtggc ggggggtaaa ctgggaaagt gatgtcgtgt 5100
actggctccg cctttttccc gaggggtggg gagaaacgta tataagtga gtagtcgccg 5160
tgaacgttct ttttcgcaac ggggtttgccc ccagaacaca gctgaagctt cgaggggctc 5220
gcattctctc ttcacgcgcg cgccgcctta cctgaggccg ccacccacgc cggttgagtc 5280
gcgttctgcc gcctcccgcc tgtggtgcct cctgaactgc gtccgcgctc taggtaagtt 5340
taaagctcag gtcgagaccg ggcctttgtc cggcgctccc ttggagccta cctagactca 5400
gcgggctctc cagcgtttgc ctgaccctgc ttgctcaact ctacgtcttt gtttcgtttt 5460
ctgtttctgc ccgttacaga tccaagctgt gacctgcgcc tacgtaagtg atatctacta 5520
gatttatcaa aaagagtgtt gacttgtgag gcctcacaaat tgatacttag attcatcgag 5580
agggacacgt cgactactaa ccttcttctc tttcctacag ctgagat 5627

```

<210> 28
 <211> 553
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pMG plasmid from InvivoGen: EMCV IRES sequence

```

<400> 28
aacgttactg gccgaagccg cttggaataa ggccgggtgt cgtttgtcta tatgttattt 60
tocaccatat tgccgtcttt tggcaatgtg agggcccggg aacctggccc tgtcttcttg 120
acgagcattc ctagggtgtt ttccctctc gccaaaggaa tgcaaggctt gttgaatgtc 180
gtgaaggaag cagttcctct ggaagcttct tgaagacaaa caacgtctgt agcgacctt 240
tgcaggcagc ggaacccccc acctggcgac aggtgcctct gcggccaaaa gccacgtgta 300
taagatacac ctgcaaaggc ggcacaaccc cagtgcacg ttgtgagttg gatagttgtg 360
gaaagagtca aatggctctc ctcaagcgtt ttcaacaagg ggctgaagga tgcccagaag 420
gtaccccatg gtatgggatc tgatctgggg cctcgggtgca catgctttac gtgtgtttag 480
tcgaggttaa aaaacgtcta ggccccccga accacgggga cgtgggtttc ctttgaaaaa 540
cacgatgata ata 553

```

<210> 29
 <211> 4692
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pDSred1-N1 plasmid from Clontech

```

<400> 29
tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg 60

```

-15-

cggtacataa	cttacggtaa	atggcccggc	tggtgacccg	cccaacgacc	cccgcccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacgg	aaactgcca	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcgggttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggacttttcca	aaatgtcgta	acaactccgc	ccatttgacg	caaatgggcg	gtaggcggtg	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaa	cgtcagatcc	gctagcgcta	600
cgggactcag	atctcgagct	caagcttcga	attctgcagt	cgacgggtacc	gcggggcccg	660
gatccaccgg	tcgccaccat	ggtgctgctc	tccaagaacg	tcatacaagg	gttcatgcgc	720
ttcaaggtgc	gcatggaggg	caccgtgaac	ggccacgagt	tcgagatcga	gggcgagggc	780
gagggccggc	cctacgaggg	ccacaacacc	gtgaagctga	aggtgaccaa	gggcgggccc	840
ctgccccttg	cctgggacat	cctgtccccc	cagttccagt	acgggtccaa	gggtgtacgtg	900
aagcaccocg	ccgacatccc	cgactacaag	aagctgtcct	tccccgaggg	cttcaagtgg	960
gagcgcgta	tgaacttcga	ggacggcggc	gtggtgaccc	tgacccagga	ctcctccctg	1020
caggacgtgc	atctcatcta	caaggtgaag	ttcacctggc	tgaacttccc	ctccgacggc	1080
cccgtaatgc	agaagaagac	catgggctgg	gagggcctca	ccgagcgcc	gtacccccgc	1140
gacggcgctg	tgaagggcga	gatccacaag	gcccgtgaag	tgaaggacgg	cgggccactac	1200
ctggtggagt	tcaagtccat	ctacatggc	aagaagccc	tcagctgcc	cggtactac	1260
tacgtggact	ccaagctgga	catcacctcc	cacaacgagg	actacacccat	cgtggagcag	1320
tacgagcgca	ccgagggccg	ccaccacctg	ttcctgtagc	ggcgcgcgact	ctagatcata	1380
atcagccata	ccacatttgt	agaggtttta	cttgccttaa	aaaacctccc	acacctcccc	1440
ctgaacctga	aacataaaa	gaatgcaatt	gttctgttta	acttgtttat	tgcagcttat	1500
aatgggtaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttcaactg	1560
cattctagtt	gtgggttgtc	caaactcatc	aatgtatcct	aaggcgtaaa	ttgtaagcgt	1620
taatattttg	ttaaaattcg	cgttaaaatt	ttgttaaatc	agctcatttt	ttaaccaata	1680
ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	accgagatag	ggttgagtgt	1740
tggtccagtt	tggaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaaggcgc	1800
aaaaacggtc	tatcaggggc	atggcccact	acgtgaacca	tcacctaat	caagtttttt	1860
ggggtcgagg	tgccgtaaa	cactaaatcg	gaaccttaaa	gggagccccc	gatttagagc	1920
ttgacgggga	aagccggcga	acgtggcgag	aaaggaagg	aagaaagcga	aaggagcggg	1980
cgctagggcg	ctggcaagtg	tagcggctac	gctgcgcgta	accaccacac	ccgcccgcgt	2040
taatggcgct	ctgacggggc	gctcaggtgg	cacttttcgg	ggaaatgtgc	gcggaaaccc	2100
tatttgttta	tttttctaaa	tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	2160
ataaatgctt	caataatatt	gaaaaaggaa	gagtcctgag	gcggaaagaa	ccagctgtgg	2220
aatgtgtgtc	agttaggggtg	tggaaagtcc	ccaggctccc	cagcaggcag	aagtatgcaa	2280
agcatgcac	tcaattagtc	agcaaccagg	tgtgaaaagt	cccaggctc	cccagcaggc	2340
agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	tagtcccggc	cctaactccg	2400
cccatcccg	ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	ctgactaatt	2460
tttttattt	atcgagaggc	cgaggccg	tcggcctctg	agctattcca	gaagttagta	2520
ggaggccttt	ttggaggcct	aggcttttgc	aaagatcgat	caagagacag	gatgaggatc	2580
gtttcgcatg	attgaacaag	atggattgca	cgcaggttct	ccggccgctt	gggtggagag	2640
gctattccgc	tatgactggg	cacaacagac	aatcggctgc	tctgatgccg	ccgtgttccg	2700
gctgtcagcg	caggggcg	cggttctttt	tgtcaagacc	gacctgtccg	gtgccctgaa	2760
tgaactgcaa	gacgaggcag	cgcggtatc	gtggctggcc	acgacggg	ttccttgcgc	2820
agctgtgctc	gacgttgtca	ctgaagcggg	aagggaactg	ctgctattgg	gcgaagtgcc	2880
ggggcaggat	ctcctgtcat	ctcaccttgc	tcctgccgag	aaagtatcca	tcattggctga	2940
tgcaatgcgg	cggtctcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	3000
acatcgcatc	gagcgagcac	gtactcggat	ggaagccggt	cttgtcgatc	aggatgatct	3060
ggacgaagag	catcaggggc	tcgcgccagc	cgaactgttc	gccaggctca	aggcgagcat	3120
gcccagcggc	gaggatctcg	tcgtgaccca	tggcgatgcc	tgcttgccga	atatcatggg	3180
ggaaaatggc	cgcttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaccgcta	3240
tcaggacata	gcgttggcta	cccgtgatat	tgctgaagag	cttggcggcg	aatgggctga	3300
ccgcttcttc	gtgcttttacg	gtatcgccgc	tcggcttcg	cagcgcatcg	ccttctatcg	3360
ccttcttgac	gagttcttct	gagcgggact	ctgggttctg	aaatgaccga	ccaagcgacg	3420
cccaacctgc	catcacgaga	tttcgattcc	accgccgcct	tctatgaaag	gttgggcttc	3480
ggaatcgctt	tccgggacgc	cggttgatg	atcctccagc	gcggggatct	catgctggag	3540
ttcttcgccc	accctagggg	gaggctaa	gaaacacgga	aggagacaat	accggaagga	3600
accgcgcgta	tgacggcaat	aaaaagacag	aataaaacgc	acggtgttgg	gtcgtttgtt	3660
cataaaacgcg	gggttcggctc	ccagggctgg	cactctgtcg	atacccacc	gagaccccat	3720
tggggccaat	acgcccgcgt	ttcttccctt	tcccccccc	accccccaag	ttcgggtgaa	3780
ggcccagggc	tcgcagccaa	cgtcggggcg	gcaggccctg	ccatagcctc	aggttactca	3840
tatatacttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	3900
ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	3960
gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	cgtaacttgc	4020
tgcttgcaaa	caaaaaaacc	accgctacca	gcggtgggtt	gtttgcccga	tcaagagcta	4080

-16-

ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	4140
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	4200
gctctgctaa	tctgtttacc	agtggctgct	gccagtggcg	ataagtcgtg	tcttaccggg	4260
ttggactcaa	gacgatagtt	accggataag	gcgagcgggt	cgggctgaac	gggggggttcg	4320
tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	4380
ctatgagaaa	gcgccacgct	tcccgaaggg	agaaaggcgg	acaggatatcc	ggtaagcggc	4440
agggtcggaa	caggagagcg	cacgaggag	cttccagggg	gaaacgcctg	gtatctttat	4500
agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	4560
gggaggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	ggccttttgc	4620
tggccttttg	ctcacatggt	ctttcctgcg	ttatccctg	attctgtgga	taaccgtatt	4680
accgccatgc	at					4692

<210> 30

<211> 4257

<212> DNA

<213> Artificial Sequence

<220>

<223> pPur plasmid from Clontech

<400> 30

ctgtggaatg	tgtgtcagtt	aggggtgtgga	aagtccccag	gctccccagc	aggcagaagt	60
atgcaaagca	tgcattctcaa	ttagttagca	accagggtgtg	gaaagtcccc	aggctcccca	120
gcaggcagaa	gtatgcaaaag	catgcatctc	aattagtcag	caaccatagt	cccgcccta	180
actccgcccc	tcccgccctt	aactccgccc	agttccgccc	attctccgcc	ccatggctga	240
ctaatttttt	ttattttatgc	agaggccgag	gcccgcctcg	cctctgagct	attccagaag	300
tagtgaggag	gctttttttg	aggcctaggc	ttttgcaaaa	agcttgcatg	cctgcaggtc	360
ggcgccacg	accgggtgct	ccaccatccc	ctgaccacac	cccctgacct	ctcacaaggga	420
gacgaccttc	catgaccgag	tacaagcca	cgggtgcgct	cgccaccgcg	gacgacgtcc	480
ccggggccgt	acgcaccctc	gccgcgcgt	tcgcccacta	ccccgccacg	cgccacaccg	540
tcgaccggga	ccgcccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	600
tcgggctcga	catcggaag	gtgtgggtcg	cggacgacgg	cgccgcgggtg	gcggctctgga	660
ccacgcggga	gagcgtcgaa	gcggggggcg	tggtcgccga	gatcgggccg	cgcatggccg	720
agttgagcgg	ttcccggtcg	gccgcgcagc	aacagatgga	aggcctcctg	gcgcgcgacc	780
ggcccaagga	gcccgcgtgg	ttcctggcca	ccgtcggcgt	ctcgcccgac	caccagcgca	840
aggggtctgg	cagcgccgtc	gtgctccccg	gagtgagggc	ggccgagcgc	gccgggggtgc	900
ccgccttctt	ggagacctcc	gcgccccgca	acctccccct	ctacgagcgg	ctcggttcca	960
ccgtcaccgc	cgagctcgag	gtgcccgaag	gaccgcgcac	ctgggtgcatg	accgcgaagc	1020
ccgggtgctg	acgcccgcgc	cacgaccgcg	agcgcccgac	cgaaaggagc	gcacgacccc	1080
atgggtccga	ccgaagccga	cccggggcgg	cccgccgacc	ccgcaccgcg	ccccgaggcc	1140
caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1200
aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tggtgtgtgt	1260
aacttgttta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	1320
aataaagcat	ttttttcact	gcattctagt	tgtgggttgt	ccaaactcat	caatgtatct	1380
tatcatgtct	ggatccccag	gaagctcctc	tggtgcctca	taaaccttaa	cctcctctac	1440
ttgagaggac	attccaatca	taggtctgcc	atccaccttc	tgtgtcctcc	tgtaatttag	1500
gtcacttaac	aaaaaggaaa	ttgggtaggg	gtttttcaca	gaccgctttc	taagggtaat	1560
tttaaaatat	ctgggaagtc	ccttccactg	ctgtgttcca	gaagtgttgg	taaacagccc	1620
acaaatgtca	acagcagaaa	catacaagct	gtcagctttg	cacaaggggc	caacacctg	1680
ctcatcaaga	agcactgtgg	ttgctgtgtt	agtaatgtgc	aaaacaggag	gcacattttc	1740
cccacctgtg	taggtttcaa	aatatctagt	gttttcattt	ttacttggat	caggaaccca	1800
gcactccact	ggataagcat	tatccttata	caaaaacagc	ttgtggtcag	tggtcatctg	1860
ctgactgtca	actgtagcat	tttttggggg	tacagtttga	gcaggatatt	tggtcctgta	1920
gtttgtctaac	acaccttgca	gctccaaagg	ttccccacca	acagcaaaaa	aatgaaaatt	1980
tgacctctga	atgggttttc	cagcaccatt	ttcatgagtt	ttttgtgtcc	ctgaatgcaa	2040
gtttaacata	gcagttaccc	caataacctc	agtttttaaca	gtaacagctt	cccacatcaa	2100
aatattttcca	cagggttaagt	cctcatttaa	attaggcaaaa	ggaattcttg	aagacgaaag	2160
ggcctcgtga	tacgcctatt	tttatagggt	aatgtcatga	taataatggg	ttcttagacg	2220
tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccctta	tttggtttatt	tttctaata	2280
cattcaaata	tgtatccgct	catgagacaa	taacctgat	aaatgcttca	ataatattga	2340
aaaaggaaga	gtatagtagt	tcaacatttc	cgtgtgcgcc	ttattccctt	ttttgctggca	2400
ttttgccttc	ctgttttttc	tcaccagaga	acgctggtga	aagtaaaaga	tgctgaagat	2460
cagttgggtg	cacgagtggt	ttacatcgaa	ctgtagctca	acagcggtaa	gatccttgag	2520
agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	ttaaagttct	gctatgtggc	2580
gcgggtattat	cccggtgtga	cgccggggcaa	gagcaactcg	gtcgccgcat	acactattct	2640
cagaatgact	tggttagtag	ctcaccagtc	acagaaaagc	atcttacgga	tggtacgaca	2700
gtaagagaat	tatgcagtgc	tgccataacc	atgagtata	acactgcggc	caacttactt	2760

-17-

ctgacaacga	tccgaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	2820
gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	2880
gacaccacga	tgccctgcag	aatggcaaca	acgttgcgca	aactattaac	tggcgaaacta	2940
cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	agttgcagga	3000
ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaaatc	tggagccggt	3060
gagcgtgggt	ctcgcgggat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	3120
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	3180
gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	3240
cttttagattg	atttaaaact	tcatttttaa	tttaaaagga	tctagggtgaa	gacccctttt	3300
gataatctca	tgacccaaaat	cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	3360
gtagaaaaga	tcaaaggatc	ttcttgagat	cctttttttc	tgccgcta	ctgctgcttg	3420
caaaacaaaa	aaccaccgct	accagcgggt	gtttgtttgc	cggatcaaga	gctaccaact	3480
ctttttccga	aggttaactg	cttcagcaga	gcgcagatac	caaatactgt	ccttctagt	3540
tagccgtagt	taggccacca	cttcaagaac	tctgtagcac	cgccacata	cctcgctctg	3600
ctaactcctg	taccagtggc	tgctgccagt	ggcgataagt	cgtgtcttac	cgggttggac	3660
tcaagacgat	agttaccgga	taaggcgag	cggctcgggt	gaacgggggg	ttcgtgcaca	3720
cagcccagct	tggaagcgaac	gacctacacc	gaactgagat	acctacagcg	tgagctatga	3780
gaaagcgcca	cgcttccga	agggagaaaag	gcgacaggt	atccggtaag	cggcaggggtc	3840
ggaacaggag	agcgacagag	ggagcttcca	gggggaaacg	cctggatatct	ttatagtcct	3900
gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	agggggggcgg	3960
agcctatgga	aaaacgccag	caacgcggcc	tttttacgggt	tcctggcctt	ttgctggcct	4020
tttgctcaca	tggtctttcc	tgcgttatcc	cctgattctg	tgataaaccg	tattaccgcc	4080
tttgagttag	ctgataccgc	tcgcccagc	cgaacagccg	agcgcagcga	gtcagtgagc	4140
gaggaagcgg	aagagcgct	gatgcggtat	tttctcctta	cgcactgttg	cggattttca	4200
caccgcata	ggtgcactct	cagtacaatc	tgctctgatg	ccgcatagtt	aagccag	4257

<210> 31
<211> 8136
<212> DNA
<213> Artificial Sequence

<220>
<223> pWE15 cosmid vector

<300>
<308> GenBank X65279
<309> 1995-04-14

<400> 31						
ctatagttag	tcgtattatg	cggccgcgaa	ttcttgaaga	cgaaagggcc	tcgtgatacg	60
ctatatttta	taggttaatg	tcattgataat	aatgggtttc	tagacgtcag	gtggcacttt	120
tcggggaaat	gtgcgcggaa	cccctatttg	tttatttttc	taaatacatt	caaataatgta	180
tcgcgtcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	240
gagtattcaa	catttccgtg	tcgcccttat	tccttttttt	gcggcatttt	gcttccgtgt	300
ttgtctcacc	cagaaaacgt	ggtgaaaagta	aaagatgctg	aagatcagtt	gggtgcacga	360
gtgggtttaca	tcgaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tcgcccggaa	420
gaacgttttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	480
gttgacgcgc	ggcaagagca	actcggtcgc	cgcatacact	attctcagaa	tgacttggtt	540
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	600
agtgtgccca	taaccatgag	tgataaacct	gcgcccaact	tacttctgac	aacgatcgga	660
ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tcgccttgat	720
cgttgggaac	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	780
gcagcaatgg	caacaacggt	gcgcaaaacta	ttactggtcg	aactacttac	tctagcttcc	840
cggcaacaat	taatagactg	gatggaggcg	gataaagttg	caggaccact	tctgcgctcg	900
gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	tggttctcgc	960
gggtatcattg	cagcactggg	gccagatggg	aagccctccc	gtatcgtagt	tatctacacg	1020
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	1080
ctgattaaagc	attggtaact	gtcagaccaa	gtttactcat	atatacttta	gattgattta	1140
aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	1200
aaaatccctt	aacgtgagtt	ttcgttccac	tgagcgtcag	accccgtaga	aaagatcaaa	1260
cagatcttctt	gagatccttt	ttttctgcgc	gtaactctgt	gcttgcaaac	aaaaaaacca	1320
ccgctaccag	cggctgggtt	tttgccggat	caagagctac	caactctttt	tccgaaggta	1380
actggcttca	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	gtagttaggc	1440
caccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgcta	cctgttacca	1500
ctggctgtgt	ccagtggcga	taagtctgtg	cttaccgggt	tggactcaag	acgatagtta	1560
cgggataagg	cgcagcggtc	gggtgaacg	gggggttcgt	gcacacagcc	cagcttgag	1620
cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaag	cgccacgctt	1680

-18-

ccgaagggag	aaaggcggac	aggatatccgg	taagcggcag	ggcgggaaca	ggagagcgca	1740
cgagggagct	tcagggggga	aacgcctggg	atcctttatag	tcctgtcggg	gtttcgccac	1800
ctctggtatg	agcgtcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	1860
gccagcaacg	cggccttttt	acggttccctg	gccttttgct	ggccttttgc	tcacatgttc	1920
tttctgcgt	tatcccttga	ttctgtggat	aaccgtatta	ccgcctttga	gtgagctgat	1980
accgctcgcc	gcagccgaac	gaccgagcgc	agcgagtcag	tgagcgagga	agcgggaagag	2040
cgctgacttc	cgcgtttcca	gactttacga	aacacggaaa	ccgaagacca	ttcatgttgt	2100
tgctcagggtc	gcagacgttt	tgacgacgca	gtcgcttcac	gttcgctcgc	gtatcggtga	2160
ttcattctgc	taaccagtaa	ggcaaccccc	ccagcctagc	cgggtcctca	acgacaggag	2220
cacgatcatg	cgcacccgtc	agatccagac	atgataagat	acattgatga	gtttggacaa	2280
accacaacta	gaatgcagtg	aaaaaaatgc	tttatttggtg	aaatttgtga	tgctattgct	2340
ttattttgtaa	ccattataag	ctgcaataaaa	caagttaaca	acaacaattg	cattcatttt	2400
atgttttcagg	ttcaggggga	ggtgtgggag	gttttttaaa	gcaagtataaa	cctctacaaa	2460
tgtgggtatgg	ctgattatga	tctctagtca	aggcactata	catcaaatat	tccttattaa	2520
cccttttaca	aattaaaaag	ctaaagggtac	acaatttttg	agcatagtta	ttaatagcag	2580
acactctatg	cctgtgtgga	gtaagaaaaa	acagtatggt	atgattataa	ctgtttgccc	2640
tacttataaa	ggttacagaa	tattttttcca	taattttctt	gtatagcagt	gcagcttttt	2700
cctttgtggg	gtaaatagca	aagcaagcaa	gagttctatt	actaaacaca	gcattgactca	2760
aaaaacttag	caattctgaa	ggaaagtcct	tggggtcttc	tacctttctc	ttcttttttg	2820
gaggagttag	atgtttgagag	tcagcagtag	cctctcatc	actagatggc	atttctctcg	2880
agcaaaaacag	gttttctca	ttaaaggcat	tcaccactcg	ctccattca	tcagttccat	2940
aggttggaat	ctaaaatata	caaacaatta	gaatcagtag	tttaacacat	tatacactta	3000
aaaattttat	atttacctta	gagctttaaa	tctctgtagg	tagtttgtcc	aattatgtca	3060
caccacagaa	gtaagggtcc	ttcacaaaaga	tcgggaccaa	agcggccatc	gtgcctcccc	3120
actcctgcag	ttcgggggca	tggatgcgcg	gatagccgct	gctggttcc	tggatgccga	3180
cggatttgca	ctgcgcgtag	aactcgcgag	gtcgccagc	ctcaggcagc	agctgaacca	3240
actcgcgagg	ggatcgagcc	cggggtgggc	gaagaactcc	agcatgagat	ccccgcgctg	3300
gaggatcatc	cagccggcgt	cccggaaaac	gattccgaag	cccaaccttt	catagaaggc	3360
ggcgttgga	tcgaaatctc	gtgatggcag	gttgggcgtc	gcttggtegg	tcatttcgaa	3420
ccccagagtc	ccgctcagaa	gaactcgtca	agaaggcgat	agaaggcgat	gcgctgcgaa	3480
tcgggagcgg	cgataccgta	aagcgcgtag	aagcggtcag	cccattcgcc	gccaaagctc	3540
tcagcaatat	cacgggtagc	caacgctatg	tcctgatagc	ggtcggccac	accagccgg	3600
ccacagtcga	tgaatccaga	aaagcggcca	ttttccacca	tgatattcgg	caagcaggca	3660
tcgcatggg	tcacgacgag	atcctcgccg	tcggggtcgc	cgcttgagc	ctggcgaaaca	3720
gttcggctgg	cgcgagcccc	tgatgctctt	cgtccagatc	atcctgatcg	acaagaccgg	3780
cttccatccg	agtacgtgct	cgtcgtatgc	gatgtttcgc	ttgggtggtcg	aatgggcagg	3840
tagccggatc	aagcgtatgc	agccgcgcga	ttgcactcagc	catgatggat	actttctcgg	3900
caggagcaag	gtgagatgac	gtgcggcagc	gcccggcagc	ttcgcccaat	agcagccagt	3960
cocttcccgc	ttcagtgaca	acgtcgagca	cagctgcgca	aggaacgccc	gtcgtggcca	4020
gccacgatag	ccgcgttgcc	tcgtcctgca	gttcattcag	ggcaccggac	aggtcgggtc	4080
tgacaaaaag	aaccggcgcc	ccttgcgctg	acagccggaa	cacggcgcca	tcagagcagc	4140
cgattgtctg	ttgtgcccag	tcatagccga	atagcctctc	cacccaagcg	gccggagaa	4200
ctgcgtgcaa	tcctcttgt	tcacatcatgc	gaacgatcc	tcactctgtc	tccttgatcag	4260
atcttgatcc	ctcgcgccat	cagatccttg	gcggcaagaa	agccatccag	tttactttgc	4320
aggccttccc	aaccttacca	gagggcgccc	cagctggcaa	ttccggttcg	cttgctgtcc	4380
ataaaaccgc	ccagtcctagc	tatcgccatg	taagcccaact	gcaagctacc	tgctttctct	4440
ttgcgcttgc	gttttccctt	gtccagatag	cccagtagct	gacattcatc	cggggtcagc	4500
accgtttctg	cggactggct	ttctacgtgt	tcgccttcc	ttagcagccc	ttgcgccctg	4560
agtgccttgc	gcagcgtgaa	agctttttgc	aaaagcctag	gcctccaaaa	aagcctcctc	4620
actacttotg	gaatagctca	gaggccgagg	cggcctaaat	aaaaaaatt	agtcagccat	4680
ggggcggaga	atgggcggaa	ctgggcggag	ttagggcgcg	gatgggcgga	gttagggcg	4740
ggactatggg	tgctgactaa	ttgagatgca	tgctttgcat	acttctgcct	gctggggagc	4800
ctggggactt	tcacacctg	gttgctgact	aatttgatag	catgctttgc	atacttctgc	4860
ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	agccggatct	4920
gcaggaccca	acgtgcccgc	agatgcgccg	cgtgcggctg	ctggagatgg	cggacgcgat	4980
ggatatgttc	tgccaagggt	tggtttgctg	attcacagtt	ctccgcaaga	attgattggc	5040
tccaattctt	ggagtgggtg	atccgcttagc	gaggtgcccgc	cggcttccat	tcaggtcgag	5100
gtggcccgcc	tcctgcacc	gacgacgaac	gcgggcaggc	agacaaggta	tagggcgccg	5160
cctacaatcc	atgccaaacc	gttccatgtg	ctcgcggagg	cgcataaatc	gccgtgacga	5220
tcagcgggtc	aatgatcgaa	gttaggctgg	taagagccgc	gagcgatcct	tgaagctgtc	5280
cctgatgggtc	gtcatctacc	tgccctggaca	gcattggcctg	caacgcggca	tcccgatgcc	5340
gcgggaagcg	agaagaatca	taatggggaa	ggccatccag	cctcgcgtcg	cgaacgccag	5400
caagacgtag	cccagcgctg	cgggcccga	tgccggcgat	aatggcctgc	ttctcgccga	5460
aacgtttggt	ggcgggacca	gtgacgaagg	cttgagcgag	ggcgtgcaag	attccgaata	5520
cgcgaagcga	caggccgac	atcgtcgccg	tcagcgaaa	cgggtcctcg	ccgaaaaatga	5580
cccagagcgc	tgccggcacc	tgtcctacga	gttgcatgat	aaagaagaca	gtcataagtg	5640
cggcgacgat	agtcatgccc	cgcgccacc	ggaaggagct	gactgggttg	aaggctctca	5700

-19-

agggcctcgg	tgcacgctct	cccttatgcg	actcctgcat	taggaagcag	cccagtagta	5760
gggtgagggc	gttgagcacc	gcccgcgcaa	ggaatgggtgc	atgcaaggag	atggcgccca	5820
acagtccccc	ggccacgggc	ctgccaccaa	acccacggcg	aaacaaggcg	tcatgagccc	5880
gaagtggcga	gcccgatcct	ccccatcggt	gatgtcggcg	atataggcgc	cagcaaccgc	5940
acctgtggcg	cgggtgatgc	cggccacgat	gcgctcggcg	tagaggatct	tggcagtcac	6000
agcatgcgca	tatccatgct	tgcacccatgc	gctcacaaaag	taggtgaatg	cgcaatgtag	6060
taccacatc	gtcatcgctt	tccactgctc	tgcgaataaa	agatggaaaa	tcaatctcat	6120
ggtaatatgc	catgaaaatc	cttgtattca	taaatcctcc	aggtagctat	atgcaaattg	6180
aaacaaaaga	gatggtgatc	tttctaagag	atgatggaat	ctcccttcag	tatcccgatg	6240
gtcaatgcgc	tggatatggg	atagatggga	atatctgat	ttttatggga	cagagttgcg	6300
aactgttccc	aactaaaatc	attttgcacg	atcagcgcac	tacgaacttt	acccacaaat	6360
agtcaggtaa	tgaatcctga	tataaagaca	ggttgataaa	tcagtcttct	acgcgcacgc	6420
cacgcgcaca	ccgtagaaaag	tctttcagtt	gtgagcctgg	gcaaaccggt	aactttcggc	6480
ggctttgctg	tgcgacaggc	tcacgtctaa	aaggaaataa	atcatgggtc	ataaaattat	6540
cacgttgtcc	ggcggcgga	cggatgttct	gtatgcgctg	ttttccgtg	gcgcgttgct	6600
gtctggtgat	ctgccttcta	aatctggcac	agccgaattg	cgcgagcttg	gttttgctga	6660
aaccagacga	acagcaactg	aataccagaa	agaaaaacac	tttacctttc	tgacatcaga	6720
agggcagaaa	tttgccgttg	aacacctggt	caatacgcgt	tttggtgagc	agcaatattg	6780
gcgtctcgatg	acgcttggcg	ttgagattga	tacctctgct	gcacaaaagg	caatcgacga	6840
ctgggaccag	cgcattcgctg	acaccgtctc	cttcgaactt	attcgcaatg	gagtgctcatt	6900
catcaaggac	gccgctatcg	caaagtgtgc	tatccacgca	gcggcaatcg	aaacacctca	6960
gccggtgacc	aatatctaca	acatcagcct	tggtatccag	cgtgatgagc	cagcgcagaa	7020
caaggttaacc	gtcagtgccg	ataagtcaa	agttaaacct	ggtgttgata	ccaacattga	7080
aacgttgatc	aaaaacgcgc	tgaaaaacgc	tgctgaatgt	gcggcgctgg	atgtcacaaa	7140
gcaaatggca	gcagacaaga	aagcgatgga	tgaactggct	tcctatgtcc	gcacggccat	7200
catgatggaa	tgtttccccg	gtggtgttat	ctggcagcag	tgccgtcgat	agtatgcaat	7260
tgataattat	tatcatttgc	gggtcctttc	cggcgatccg	ccttgttacg	ggcgggcgac	7320
ctcgcgggtt	ttcgctattt	atgaaaattt	tccggtttta	ggcgtttccg	ttcttcttcg	7380
tcataactta	atgtttttat	ttaaaatacc	ctctgaaaag	aaaggaaaacg	acaggtgctg	7440
aaagcgagct	ttttggcctc	tgctgtttcc	tttctctgtt	tttgtccgtg	gaatgaacaa	7500
tggaagtcaa	caaaaagcag	ctgggtgaca	ttttcggtgc	gagtatccgt	accattcaga	7560
actggcagga	acagggaatg	cccgttctgc	gaggcggtgg	caagggtaat	gaggtgcttt	7620
atgactctgc	cgcgctcata	aaatggtatg	ccgaaaggga	tgctgaaatt	gagaacgaaa	7680
agctgcgccg	ggaggttgaa	gaactgcggc	aggccagcga	ggcagatcca	caggacgggt	7740
gtggtcgcca	tgatcgcgta	gtcgatagtg	gctccaagta	gcgaagcgag	caggactggg	7800
cggcggaaca	gcggtcggac	agtgtccga	gaacgggtgc	gcatagaaat	tgcatcaacg	7860
catatagcgc	tagcagcacg	ccatagtgc	tgccgatgct	gtcggaatgg	acgatattcc	7920
gcaagaggcc	ggcataacca	ggcataacca	agccatagcc	tacagcatcc	aggggtgacgg	7980
tgccgaggat	gacgatgagc	gcattgttag	atttcataca	cggtgccctga	ctgcgttagc	8040
aatttaactg	tgataaacta	cgcattaaaa	gcttatcgat	gataagcggg	caaacatgag	8100
aattcgcggc	cgcaattaac	cctcactaaa	ggatcc			8136

<210> 32
 <211> 2713
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pNEB193 plasmid

<400> 32						
tgcgcggttt	cggatgatgac	gggtgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gcccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	cgcacagat	gcgtaaaggag	aaaataccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aactgttggg	aaggcgatc	ggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaaagg	ggatgtgctg	caaggcgatt	aagttgggtg	acgccagggg	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgaatt	cgagctcggg	acccgggggc	420
gcgcgggatc	cttaattaa	tctagagtcg	actgtttaaa	cctgcaggca	tgcaagcttg	480
gcgtaatcat	ggtcatagct	gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	540
aacatacgag	ccggaagcat	aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaaactc	600
acattaattg	cgttgcgctc	actgccgctc	ttccagtcgg	gaaacctgtc	gtgccagctg	660
cattaatgaa	tcggccaacg	cgcggggaga	ggcgttttgc	gtattgggcg	ctcttcgcgt	720
tcttcgctca	ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggg	atcagctcac	780
tcaaaaggcg	taatacgggt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	840
gcaaaaggcc	agcaaaaggc	caggaaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	900
aggctccgcc	cccctgacga	gcatcacaaa	aatcgacgct	caagtcagag	gtggcgaaac	960

-20-

ccgacaggac	tataaagata	ccaggcggtt	ccccctggaa	gctccctcgt	gcgctctcct	1020
gttccgaccc	tgccgcttac	cggataacct	tccgcctttc	tcccttcggg	aagcgtggcg	1080
ctttctcata	gctcacgctg	taggtatctc	agttcgggtg	aggctcgttcg	ctccaagctg	1140
ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	ccttatcccg	taactatcgt	1200
cttgagtcca	acccggttaag	acacgactta	tgcgactg	cagcagccac	tggtaacagg	1260
attagcagag	cgaggatatgt	aggcgggtgt	acagagttct	tgaagtgggtg	gcctaactac	1320
ggctacacta	gaaggacagt	atlttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	1380
aaaagagttg	gtagctccttg	atccggcaaaa	caaaccaccg	ctggttagcgg	tggttttttt	1440
gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	1500
tctacggggt	ctgacgctca	gtggaaacgaa	aactcacggt	aagggatttt	ggcatgaga	1560
ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	1620
taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	gcttaatcag	tgaggcacct	1680
atctcagcga	tctgtctatt	tgcgtcatcc	atagttgcct	gactccccgt	cgtgtagata	1740
actacgatac	gggagggtct	accatctggc	cccagtgctg	caatgatacc	gcgagaccca	1800
cgctcaccgg	ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	1860
agtggctcct	caactttatc	cgctccatc	cagtctatta	attgttgccg	ggaagctaga	1920
gtaagtagtt	cgccagttaa	tagtttgccg	aacggtgttg	ccattgctac	aggcatcgtg	1980
gtgtcacgct	cgctcgtttg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	2040
gttacatgat	cccccatgtt	gtgcaaaaaa	gcggttagct	ccttcggtcc	tccgatcgtt	2100
gtcagaagta	agttggccgc	agtgttatca	ctcatgggta	tggcagcact	gcataattct	2160
cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	2220
ttctgagaat	agtgtatgcg	gcgaccgagt	tgcctctg	cggcgtcaat	acgggataat	2280
accgcgccac	atagcagaac	tttaaaagtg	ctcatcattg	gaaaacgttc	ttcggggcga	2340
aaactctcaa	ggatcttacc	gctgttgaga	tccagttcga	tgtaacccac	tcgtgcaccc	2400
aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	2460
caaaatgccg	caaaaaaggg	aataagggcg	acacggaaat	gttgaatact	catactcttc	2520
ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatattt	2580
gaatgtat	agaaaaataa	acaaataggg	gttcgcgcga	catttccccg	aaaagtgcc	2640
cctgacgtct	aagaaacat	tattatcatg	acattaacct	ataaaaaatag	gcgtatcacg	2700
aggccctttc	gtc					2713

<210> 33
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attP

<400> 33
 cagctttttt atactaagtt g 21

<210> 34
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attB

<400> 34
 ctgctttttt atactaactt g 21

<210> 35
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attL

<400> 35
 ctgctttttt atactaagtt g 21

<210> 36
 <211> 21
 <212> DNA

-21-

<213> Artificial Sequence

<220>

<223> attR

<400> 36

cagctttttt atactaactt g

21

<210> 37

<211> 1071

<212> DNA

<213> Artificial Sequence

<220>

<223> Integrase E174R

<221> CDS

<222> (1)...(1071)

<223> Nucleotide sequence encoding Integrase E147R

<400> 37

atg gga aga agg cga agt cat gag cgc cgg gat tta ccc cct aac ctt	48
Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
tat ata aga aac aat gga tat tac tgc tac agg gac cca agg acg ggt	96
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
aaa gag ttt gga tta ggc aga gac agg cga atc gca atc act gaa gct	144
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
ata cag gcc aac att gag tta ttt tca gga cac aaa cac aag cct ctg	192
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
aca gcg aga atc aac agt gat aat tcc gtt acg tta cat tca tgg ctt	240
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
gat cgc tac gaa aaa atc ctg gcc agc aga gga atc aag cag aag aca	288
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	
85 90 95	
ctc ata aat tac atg agc aaa att aaa gca ata agg agg ggt ctg cct	336
Leu Ile Asn Tyr Met Ser Lys Ile Lys Ala Ile Arg Arg Gly Leu Pro	
100 105 110	
gat gct cca ctt gaa gac atc acc aca aaa gaa att gcg gca atg ctc	384
Asp Ala Pro Leu Glu Asp Ile Thr Thr Lys Glu Ile Ala Ala Met Leu	
115 120 125	
aat gga tac ata gac gag ggc aag gcg gcg tca gcc aag tta atc aga	432
Asn Gly Tyr Ile Asp Glu Gly Lys Ala Ala Ser Ala Lys Leu Ile Arg	
130 135 140	
tca aca ctg agc gat gca ttc cga gag gca ata gct gaa ggc cat ata	480
Ser Thr Leu Ser Asp Ala Phe Arg Glu Ala Ile Ala Glu Gly His Ile	
145 150 155 160	
aca aca aac cat gtc gct gcc act cgc gca gca aaa tct aga gta agg	528
Thr Thr Asn His Val Ala Ala Thr Arg Ala Ala Lys Ser Arg Val Arg	
165 170 175	
aga tca aga ctt acg gct gac gaa tac ctg aaa att tat caa gca gca	576
Arg Ser Arg Leu Thr Ala Asp Glu Tyr Leu Lys Ile Tyr Gln Ala Ala	

-22-

180										185										190										
gaa tca tca cca tgt tgg ctc	aga ctt gca atg gaa ctg gct gtt gtt	624																												
Glu Ser Ser Pro Cys Trp Leu Arg	Leu Ala Met Glu Leu Ala Val Val																													
195	200	205																												
acc ggg caa cga gtt ggt gat tta tgc gaa atg aag tgg tct gat atc	672																													
Thr Gly Gln Arg Val Gly Asp Leu Cys Glu Met Lys Trp Ser Asp Ile																														
210	215	220																												
gta gat gga tat ctt tat gtc gag caa agc aaa aca ggc gta aaa att	720																													
Val Asp Gly Tyr Leu Tyr Val Glu Gln Ser Lys Thr Gly Val Lys Ile																														
225	230	235	240																											
gcc atc cca aca gca ttg cat att gat gct ctc gga ata tca atg aag	768																													
Ala Ile Pro Thr Ala Leu His Ile Asp Ala Leu Gly Ile Ser Met Lys																														
245	250	255																												
gaa aca ctt gat aaa tgc aaa gag att ctt ggc gga gaa acc ata att	816																													
Glu Thr Leu Asp Lys Cys Lys Glu Ile Leu Gly Gly Glu Thr Ile Ile																														
260	265	270																												
gca tct act cgt cgc gaa ccg ctt tca tcc ggc aca gta tca agg tat	864																													
Ala Ser Thr Arg Arg Glu Pro Leu Ser Ser Gly Thr Val Ser Arg Tyr																														
275	280	285																												
ttt atg cgc gca cga aaa gca tca ggt ctt tcc ttc gaa ggg gat ccg	912																													
Phe Met Arg Ala Arg Lys Ala Ser Gly Leu Ser Phe Glu Gly Asp Pro																														
290	295	300																												
cct acc ttt cac gag ttg cgc agt ttg tct gca aga ctc tat gag aag	960																													
Pro Thr Phe His Glu Leu Arg Ser Leu Ser Ala Arg Leu Tyr Glu Lys																														
305	310	315	320																											
cag ata agc gat aag ttt gct caa cat ctt ctc ggg cat aag tgc gac	1008																													
Gln Ile Ser Asp Lys Phe Ala Gln His Leu Leu Gly His Lys Ser Asp																														
325	330	335																												
acc atg gca tca cag tat cgt gat gac aga ggc agg gag tgg gac aaa	1056																													
Thr Met Ala Ser Gln Tyr Arg Asp Asp Arg Gly Arg Glu Trp Asp Lys																														
340	345	350																												
att gaa atc aaa taa	1071																													
Ile Glu Ile Lys *																														
355																														

<210> 38

<211> 356

<212> PRT

<213> Artificial Sequence

<220>

<223> Integrase B147R

<400> 38

Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	

[illegible]

```
<220>
<221> CDS
<222> (45)...(737)
<223> Nucleotide sequence encoding red fluorescent
      protein (FP593)
```

```
<400> 39
agtttcagcc agtgacaggg tgagctgcc a ggtattctaa caag atg agt tgt tcc 56
Met Ser Cys Ser
1
```

cct tac gaa ggt cac tgt tcc gta aag ctt atg gta acc aag ggt gga 200
Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val Thr Lys Gly Gly

-24-

40	45	50	
cct ttg cca ttt gct ttt gat att ttg tca cca caa ttt cag tat gga	248		
Pro Leu Pro Phe Ala Phe Asp Ile Leu Ser Pro Gln Phe Gln Tyr Gly			
55 60 65			
agc aag gta tat gtc aaa cac cct gcc gac ata cca gac tat aaa aag	296		
Ser Lys Val Tyr Val Lys His Pro Ala Asp Ile Pro Asp Tyr Lys Lys			
70 75 80			
ctg tca ttt cct gag gga ttt aaa tgg gaa agg gtc atg aac ttt gaa	344		
Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val Met Asn Phe Glu			
85 90 95 100			
gac ggt ggc gtg gtt act gta tcc caa gat tcc agt ttg aaa gac ggc	392		
Asp Gly Gly Val Val Thr Val Ser Gln Asp Ser Ser Leu Lys Asp Gly			
105 110 115			
tgt ttc atc tac gag gtc aag ttc att ggg gtg aac ttt cct tct gat	440		
Cys Phe Ile Tyr Glu Val Lys Phe Ile Gly Val Asn Phe Pro Ser Asp			
120 125 130			
gga cct gtt atg cag agg agg aca cgg ggc tgg gaa gcc agc tct gag	488		
Gly Pro Val Met Gln Arg Arg Thr Arg Gly Trp Glu Ala Ser Ser Glu			
135 140 145			
cgt ttg tat cct cgt gat ggg gtg ctg aaa gga gac atc cat atg gct	536		
Arg Leu Tyr Pro Arg Asp Gly Val Leu Lys Gly Asp Ile His Met Ala			
150 155 160			
ctg agg ctg gaa gga ggc ggc cat tac ctc gtt gaa ttc aaa agt att	584		
Leu Arg Leu Glu Gly Gly Gly His Tyr Leu Val Glu Phe Lys Ser Ile			
165 170 175 180			
tac atg gta aag aag cct tca gtg cag ttg cca ggc tac tat tat gtt	632		
Tyr Met Val Lys Lys Pro Ser Val Gln Leu Pro Gly Tyr Tyr Tyr Val			
185 190 195			
gac tcc aaa ctg gat atg acg agc cac aac gaa gat tac aca gtc gtt	680		
Asp Ser Lys Leu Asp Met Thr Ser His Asn Glu Asp Tyr Thr Val Val			
200 205 210			
gag cag tat gaa aaa acc cag gga cgc cac cat ccg ttc att aag cct	728		
Glu Gln Tyr Glu Lys Thr Gln Gly Arg His His Pro Phe Ile Lys Pro			
215 220 225			
ctg cag tga actcggctca gtcattgatt agcggtaatg gccacaaaag	777		
Leu Gln *			
230			
gcacgatgat cgtttttttag gaatgcagcc aaaaattgaa ggttatgaca gtagaaatac	837		
aagcaacagg ctttgcttat taaacatgta attgaaaac	876		

<210> 40

<211> 230

<212> PRT

<213> Discosoma species

<400> 40

Met Ser Cys Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val	
1 5 10 15	
Arg Met Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Lys Gly Glu	
20 25 30	
Gly Glu Gly Arg Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val	
35 40 45	
Thr Lys Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ser Pro Gln	
50 55 60	

-25-

```

Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His Pro Ala Asp Ile Pro
65      70      75      80
Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val
      85      90      95
Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Ser Gln Asp Ser Ser
      100      105      110
Leu Lys Asp Gly Cys Phe Ile Tyr Glu Val Lys Phe Ile Gly Val Asn
      115      120      125
Phe Pro Ser Asp Gly Pro Val Met Gln Arg Arg Thr Arg Gly Trp Glu
      130      135      140
Ala Ser Ser Glu Arg Leu Tyr Pro Arg Asp Gly Val Leu Lys Gly Asp
145      150      155      160
Ile His Met Ala Leu Arg Leu Glu Gly Gly His Tyr Leu Val Glu
      165      170      175
Phe Lys Ser Ile Tyr Met Val Lys Lys Pro Ser Val Gln Leu Pro Gly
      180      185      190
Tyr Tyr Tyr Val Asp Ser Lys Leu Asp Met Thr Ser His Asn Glu Asp
      195      200      205
Tyr Thr Val Val Glu Gln Tyr Glu Lys Thr Gln Gly Arg His His Pro
      210      215      220
Phe Ile Lys Pro Leu Gln
225      230

```

```

<210> 41
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-att;

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

```

<400> 41
rkycwgcttt yktrtacnaa stsgb

```

25

```

<210> 42
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-attB;

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

```

<400> 42
agccwgcttt yktrtacnaa ctsgb

```

25

```

<210> 43
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-attR

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

-26-

<400> 43
gttcagcttt cktrtacnaa ctsgb 25

<210> 44
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attL

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 44
agccwgcttt cktrtacnaa gtsgb 25

<210> 45
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attP1

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 45
gttcagcttt yktrtacnaa gtsgb 25

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 46
agcctgcttt tttgtacaaa cttgt 25

<210> 47
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 47
agcctgcttt cttgtacaaa cttgt 25

<210> 48
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB3

<400> 48
accagcttt cttgtacaaa cttgt 25

<210> 49

-27-

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR1

<400> 49
gttcagcttt tttgtacaaa cttgt 25

<210> 50
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR2

<400> 50
gttcagcttt cttgtacaaa cttgt 25

<210> 51
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR3

<400> 51
gttcagcttt cttgtacaaa gttgg 25

<210> 52
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL1

<400> 52
agcctgcttt tttgtacaaa gttgg 25

<210> 53
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL2

<400> 53
agcctgcttt cttgtacaaa gttgg 25

<210> 54
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL3

<400> 54
accagcttt cttgtacaaa gttgg 25

<210> 55
<211> 25

-28-

<212> DNA
 <213> Artificial Sequence

 <220>
 <223> attP1

 <400> 55
 gttcagcttt tttgtacaaa gttgg 25

 <210> 56
 <211> 25
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> attP2,P3

 <400> 56
 gttcagcttt cttgtacaaa gttgg 25

 <210> 57
 <211> 34
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Lox P site

 <400> 57
 ataacttcgt ataatgtatg ctatacgaag ttat 34

 <210> 58
 <211> 1032
 <212> DNA
 <213> Escherichia coli

 <220>
 <221> CDS
 <222> (1)...(1032)
 <223> nucleotide sequence encoding Cre recombinase

 <400> 58
 atg tcc aat tta ctg acc gta cac caa aat ttg cct gca tta ccg gtc 48
 Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
 1 5 10 15

 gat gca acg agt gat gag gtt cgc aag aac ctg atg gac atg ttc agg 96
 Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
 20 25 30

 gat cgc cag gcg ttt tct gag cat acc tgg aaa atg ctt ctg tcc gtt 144
 Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
 35 40 45

 tgc cgg tcg tgg gcg gca tgg tgc aag ttg aat aac cgg aaa tgg ttt 192
 Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
 50 55 60

 ccc gca gaa cct gaa gat gtt cgc gat tat ctt cta tat ctt cag gcg 240
 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
 65 70 75 80

 cgc ggt ctg gca gta aaa act atc cag caa cat ttg ggc cag cta aac 288
 Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
 85 90 95

 atg ctt cat cgt cgg tcc ggg ctg cca cga cca agt gac agc aat gct 336

-29-

Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	Asn	Ala		
			100					105					110				
ggt	tca	ctg	ggt	atg	cgg	cgg	atc	cga	aaa	gaa	aac	ggt	gat	gcc	ggt	384	
Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	Ala	Gly		
		115					120					125					
gaa	cgt	gca	aaa	cag	gct	cta	gcg	ttc	gaa	cgc	act	gat	ttc	gac	cag	432	
Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	Asp	Gln		
		130				135					140						
ggt	cgt	tca	ctc	atg	gaa	aat	agc	gat	cgc	tgc	cag	gat	ata	cgt	aat	480	
Val	Arg	Ser	Leu	Met	Gly	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	Arg	Asn		
145					150					155					160		
ctg	gca	ttt	ctg	ggg	att	gct	tat	aac	acc	ctg	tta	cgt	ata	gcc	gaa	528	
Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	Ala	Glu		
				165					170					175			
att	gcc	agg	atc	agg	ggt	aaa	gat	atc	tca	cgt	act	gac	ggt	ggg	aga	576	
Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	Gly	Arg		
			180					185					190				
atg	tta	atc	cat	att	ggc	aga	acg	aaa	acg	ctg	ggt	agc	acc	gca	ggt	624	
Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	Ala	Gly		
			195				200					205					
gta	gag	aag	gca	ctt	agc	ctg	ggg	gta	act	aaa	ctg	gtc	gag	cga	tgg	672	
Val	Glu	Lys	Ala	Leu	Ser	Leu	Gly	Val	Thr	Lys	Leu	Val	Glu	Arg	Trp		
		210				215					220						
att	tcc	gtc	tct	ggt	gta	gct	gat	gat	ccg	aat	aac	tac	ctg	ttt	tgc	720	
Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	Phe	Cys		
225					230					235					240		
cgg	gtc	aga	aaa	aat	ggt	ggt	gcc	gcg	cca	tct	gcc	acc	agc	cag	cta	768	
Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	Gln	Leu		
				245					250					255			
tca	act	cgc	gcc	ctg	gaa	ggg	att	ttt	gaa	gca	act	cat	cga	ttg	att	816	
Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	Leu	Ile		
			260					265					270				
tac	ggc	gct	aag	gat	gac	tct	ggt	cag	aga	tac	ctg	gcc	tgg	tct	gga	864	
Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	Ser	Gly		
		275					280					285					
cac	agt	gcc	cgt	gtc	gga	gcc	gcg	cga	gat	atg	gcc	cgc	gct	gga	ggt	912	
His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	Gly	Val		
		290				295					300						
tca	ata	ccg	gag	atc	atg	caa	gct	ggt	ggc	tgg	acc	aat	gta	aat	att	960	
Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	Asn	Ile		
305					310					315					320		
gtc	atg	aac	tat	atc	cgt	aac	ctg	gat	agt	gaa	aca	ggg	gca	atg	gtg	1008	
Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	Met	Val		
				325					330					335			
cgc	ctg	ctg	gaa	gat	ggc	gat	tag									1032	
Arg	Leu	Leu	Glu	Asp	Gly	Asp	*										
			340														

<210> 59
 <211> 343
 <212> PRT

-30-

<213> Escherichia coli

<400> 59

```

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1      5      10      15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
20      25      30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
35      40      45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
50      55      60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Tyr Leu Gln Ala
65      70      75      80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
85      90      95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
100      105      110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
115      120      125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
130      135      140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
145      150      155      160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
165      170      175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
180      185      190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
195      200      205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
210      215      220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
225      230      235      240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
245      250      255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
260      265      270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
275      280      285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
290      295      300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
305      310      315      320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
325      330      335
Arg Leu Leu Glu Asp Gly Asp
340

```

<210> 60

<211> 1272

<212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)...(1272)

<223> nucleotide sequence encoding Flip recombinase

<400> 60

```

atg cca caa ttt ggt ata tta tgt aaa aca cca cct aag gtg ctt gtt
Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1      5      10      15
cgt cag ttt gtg gaa agg ttt gaa aga cct tca ggt gag aaa ata gca
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
20      25      30

```

48

96

-31-

tta	tgt	gct	gct	gaa	cta	acc	tat	tta	tgt	tggt	atg	att	aca	cat	aac	144
Leu	Cys	Ala	Ala	Glu	Leu	Thr	Tyr	Leu	Cys	Trp	Met	Ile	Thr	His	Asn	
		35					40					45				
gga	aca	gca	atc	aag	aga	gcc	aca	ttc	atg	agc	tat	aat	act	atc	ata	192
Gly	Thr	Ala	Ile	Lys	Arg	Ala	Thr	Phe	Met	Ser	Tyr	Asn	Thr	Ile	Ile	
	50					55					60					
agc	aat	tcg	ctg	agt	ttc	gat	att	gtc	aat	aaa	tca	ctc	cag	ttt	aaa	240
Ser	Asn	Ser	Leu	Ser	Phe	Asp	Ile	Val	Asn	Lys	Ser	Leu	Gln	Phe	Lys	
65					70					75					80	
tac	aag	acg	caa	aaa	gca	aca	att	ctg	gaa	gcc	tca	tta	aag	aaa	ttg	288
Tyr	Lys	Thr	Gln	Lys	Ala	Thr	Ile	Leu	Glu	Ala	Ser	Leu	Lys	Lys	Leu	
			85						90					95		
att	cct	gct	tggt	gaa	ttt	aca	att	att	cct	tac	tat	gga	caa	aaa	cat	336
Ile	Pro	Ala	Trp	Glu	Phe	Thr	Ile	Ile	Pro	Tyr	Tyr	Gly	Gln	Lys	His	
			100					105					110			
caa	tct	gat	atc	act	gat	att	gta	agt	agt	ttg	caa	tta	cag	ttc	gaa	384
Gln	Ser	Asp	Ile	Thr	Asp	Ile	Val	Ser	Ser	Leu	Gln	Leu	Gln	Phe	Glu	
		115					120					125				
tca	tcg	gaa	gaa	gca	gat	aag	gga	aat	agc	cac	agt	aaa	aaa	atg	ctt	432
Ser	Ser	Glu	Glu	Ala	Asp	Lys	Gly	Asn	Ser	His	Ser	Lys	Lys	Met	Leu	
	130					135					140					
aaa	gca	ctt	cta	agt	gag	gggt	gaa	agc	atc	tggt	gag	atc	act	gag	aaa	480
Lys	Ala	Leu	Leu	Ser	Glu	Gly	Glu	Ser	Ile	Trp	Glu	Ile	Thr	Glu	Lys	
145					150					155					160	
ata	cta	aat	tcg	ttt	gag	tat	act	tcg	aga	ttt	aca	aaa	aca	aaa	act	528
Ile	Leu	Asn	Ser	Phe	Glu	Tyr	Thr	Ser	Arg	Phe	Thr	Lys	Thr	Lys	Thr	
				165					170					175		
tta	tac	caa	ttc	ctc	ttc	cta	gct	act	ttc	atc	aat	tgt	gga	aga	ttc	576
Leu	Tyr	Gln	Phe	Leu	Phe	Leu	Ala	Thr	Phe	Ile	Asn	Cys	Gly	Arg	Phe	
			180					185					190			
agc	gat	att	aag	aac	ggt	gat	ccg	aaa	tca	ttt	aaa	tta	gtc	caa	aat	624
Ser	Asp	Ile	Lys	Asn	Val	Asp	Pro	Lys	Ser	Phe	Lys	Leu	Val	Gln	Asn	
		195					200						205			
aag	tat	ctg	gga	gta	ata	atc	cag	tgt	tta	gtg	aca	gag	aca	aag	aca	672
Lys	Tyr	Leu	Gly	Val	Ile	Ile	Gln	Cys	Leu	Val	Thr	Glu	Thr	Lys	Thr	
	210					215					220					
agc	ggt	agt	agg	cac	ata	tac	ttc	ttt	agc	gca	agg	ggt	agg	atc	gat	720
Ser	Val	Ser	Arg	His	Ile	Tyr	Phe	Phe	Ser	Ala	Arg	Gly	Arg	Ile	Asp	
225					230					235					240	
cca	ctt	gta	tat	ttg	gat	gaa	ttt	ttg	agg	aat	tct	gaa	cca	gtc	cta	768
Pro	Leu	Val	Tyr	Leu	Asp	Glu	Phe	Leu	Arg	Asn	Ser	Glu	Pro	Val	Leu	
				245					250					255		
aaa	cga	gta	aat	agg	acc	ggc	aat	tct	tca	agc	aat	aaa	cag	gaa	tac	816
Lys	Arg	Val	Asn	Arg	Thr	Gly	Asn	Ser	Ser	Ser	Asn	Lys	Gln	Glu	Tyr	
			260					265					270			
caa	tta	tta	aaa	gat	aac	tta	gtc	aga	tcg	tac	aat	aaa	gct	ttg	aag	864
Gln	Leu	Leu	Lys	Asp	Asn	Leu	Val	Arg	Ser	Tyr	Asn	Lys	Ala	Leu	Lys	
		275					280					285				
aaa	aat	gcg	cct	tat	tca	atc	ttt	gct	ata	aaa	aat	ggc	cca	aaa	tct	912
Lys	Asn	Ala	Pro	Tyr	Ser	Ile	Phe	Ala	Ile	Lys	Asn	Gly	Pro	Lys	Ser	
	290					295					300					

-32-

cac att gga aga cat ttg atg acc tca ttt ctt tca atg aag ggc cta 960
 His Ile Gly Arg His Leu Met Thr Ser Phe Leu Ser Met Lys Gly Leu
 305 310 315 320
 acg gag ttg act aat gtt gtg gga aat tgg agc gat aag cgt gct tct 1008
 Thr Glu Leu Thr Asn Val Val Gly Asn Trp Ser Asp Lys Arg Ala Ser
 325 330 335
 gcc gtg gcc agg aca acg tat act cat cag ata aca gca ata cct gat 1056
 Ala Val Ala Arg Thr Thr Tyr Thr His Gln Ile Thr Ala Ile Pro Asp
 340 345 350
 cac tac ttc gca cta gtt tct cgg tac tat gca tat gat cca ata tca 1104
 His Tyr Phe Ala Leu Val Ser Arg Tyr Tyr Ala Tyr Asp Pro Ile Ser
 355 360 365
 aag gaa atg ata gca ttg aag gat gag act aat cca att gag gag tgg 1152
 Lys Glu Met Ile Ala Leu Lys Asp Glu Thr Asn Pro Ile Glu Glu Trp
 370 375 380
 cag cat ata gaa cag cta aag ggt agt gct gaa gga agc ata cga tac 1200
 Gln His Ile Glu Gln Leu Lys Gly Ser Ala Glu Gly Ser Ile Arg Tyr
 385 390 395 400
 ccc gca tgg aat ggg ata ata tca cag gag gta cta gac tac ctt tca 1248
 Pro Ala Trp Asn Gly Ile Ile Ser Gln Glu Val Leu Asp Tyr Leu Ser
 405 410 415
 tcc tac ata aat aga cgc ata taa 1272
 Ser Tyr Ile Asn Arg Arg Ile *
 420

<210> 61

<211> 422

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 61

Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val Arg
 1 5 10 15
 Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala Leu
 20 25 30
 Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn Gly
 35 40 45
 Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile Ser
 50 55 60
 Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys Tyr
 65 70 75 80
 Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu Ile
 85 90 95
 Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His Gln
 100 105 110
 Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu Ser
 115 120 125
 Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu Lys
 130 135 140
 Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys Ile
 145 150 155 160
 Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr Leu
 165 170 175
 Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe Ser
 180 185 190
 Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn Lys
 195 200 205
 Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr Ser

-33-

```

      210      215      220
Val Ser Arg His Ile Tyr Phe Phe Ser Ala Arg Gly Arg Ile Asp Pro
225      230      235      240
Leu Val Tyr Leu Asp Glu Phe Leu Arg Asn Ser Glu Pro Val Leu Lys
      245      250      255
Arg Val Asn Arg Thr Gly Asn Ser Ser Ser Asn Lys Gln Glu Tyr Gln
      260      265      270
Leu Leu Lys Asp Asn Leu Val Arg Ser Tyr Asn Lys Ala Leu Lys Lys
      275      280      285
Asn Ala Pro Tyr Ser Ile Phe Ala Ile Lys Asn Gly Pro Lys Ser His
      290      295      300
Ile Gly Arg His Leu Met Thr Ser Phe Leu Ser Met Lys Gly Leu Thr
305      310      315      320
Glu Leu Thr Asn Val Val Gly Asn Trp Ser Asp Lys Arg Ala Ser Ala
      325      330      335
Val Ala Arg Thr Thr Tyr Thr His Gln Ile Thr Ala Ile Pro Asp His
      340      345      350
Tyr Phe Ala Leu Val Ser Arg Tyr Tyr Ala Tyr Asp Pro Ile Ser Lys
      355      360      365
Glu Met Ile Ala Leu Lys Asp Glu Thr Asn Pro Ile Glu Glu Trp Gln
      370      375      380
His Ile Glu Gln Leu Lys Gly Ser Ala Glu Gly Ser Ile Arg Tyr Pro
385      390      395      400
Ala Trp Asn Gly Ile Ile Ser Gln Glu Val Leu Asp Tyr Leu Ser Ser
      405      410      415
Tyr Ile Asn Arg Arg Ile
      420

```

```

<210> 62
<211> 48
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> IR2

```

```

<400> 62
gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttc
48

```

```

<210> 63
<211> 48
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> IR1

```

```

<400> 63
gaagttccta tactttctag agaataggaa cttcggaata ggaacttc
48

```

```

<210> 64
<211> 66
<212> DNA
<213> Bacteriophage mu

```

```

<220>
<221> CDS
<222> (1)...(66)
<223> nucleotide sequence encoding GIN recombinase

```

```

<400> 64
tca act ctg tat aaa aaa cac ccc gcg aaa cga gcg cat ata gaa aac
Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
1 5 10 15
48

```

```

gac gat cga atc aat taa
Asp Asp Arg Ile Asn *
66

```

-34-

20

<210> 65
 <211> 21
 <212> PRT
 <213> bacteriophage mu

<400> 65
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
 1 5 10 15
 Asp Asp Arg Ile Asn
 20

<210> 66
 <211> 69
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(69)
 <223> nucleotide sequence encoding Gin recombinase

<400> 66
 tat aaa aaa cat ccc gcg aaa cga acg cat ata gaa aac gac gat cga 48
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15
 atc aat caa atc gat cgg taa 69
 Ile Asn Gln Ile Asp Arg *
 20

<210> 67
 <211> 22
 <212> PRT
 <213> bacteriophage mu

<220>
 <223> Gin recombinase of bacteriophage mu

<400> 67
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
 1 5 10 15
 Ile Asn Gln Ile Asp Arg
 20

<210> 68
 <211> 555
 <212> DNA
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)...(555)
 <223> nucleotide sequence encoding PIN recombinase

<400> 68
 atg ctt att ggc tat gta cgc gta tca aca aat gac cag aac aca gat 48
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15
 cta caa cgt aat gcg ctg aac tgt gca gga tgc gag ctg att ttt gaa 96
 Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu
 20 25 30

-35-

gac aag ata agc ggc aca aag tcc gaa agg ccg gga ctg aaa aaa ctg	144
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
ctc agg aca tta tcg gca ggt gac act ctg gtt gtc tgg aag ctg gat	192
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
cgg ctg ggg cgt agt atg cgg cat ctt gtc gtg ctg gtg gag gag ttg	240
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
cgc gaa cga ggc atc aac ttt cgt agt ctg acg gat tca att gat acc	288
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
agc aca cca atg gga cgc ttt ttc ttt cat gtg atg ggt gcc ctg gct	336
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
gaa atg gag cgt gaa ctg att gtt gaa cga aca aaa gct gga ctg gaa	384
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
act gct cgt gca cag gga cga att ggt gga cgt cgt ccc aaa ctt aca	432
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
cca gaa caa tgg gca caa gct gga cga tta att gca gca gga act cct	480
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	
cgc cag aag gtg gcg att atc tat gat gtt ggt gtg tca act ttg tat	528
Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr	
165 170 175	
aag agg ttt cct gca ggg gat aaa taa	555
Lys Arg Phe Pro Ala Gly Asp Lys *	
180	

<210> 69

<211> 184

<212> PRT

<213> Escherichia coli

<400> 69

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp	
1 5 10 15	
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu	
20 25 30	
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	

-36-

Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr
 165 170 175
 Lys Arg Phe Pro Ala Gly Asp Lys
 180

<210> 70
 <211> 4778
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pcx plasmid

<400> 70
 gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
 gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 120
 ccaacgaccc ccgccatttg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
 ggacttttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtag 240
 atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
 cctggcatta tgcccagtag atgaccttat gggactttcc tacttggcag tacatctacg 360
 tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
 atctcccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
 gcgatggggg cggggggggg gggggcgcg cccaggcggg gcggggcggg gcgagggggc 540
 gggcgggggc aggcggagag gtgcggcggg agccaatcag agcggcgcg cccgaaagtt 600
 tcctttttatg gcgagcgggc ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc 660
 gggagtcgct gcgttgccct cgcgcccggt gccgctccgc gccgccccgc gccgccccgc 720
 ccggctctga ctgaccgcgt tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780
 gggctgtaac tagcgcttgg tttaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840
 ccttaaaggc ctccgggagg gccctttgtg cgggggggag cggctcgggg ggtgctgctg 900
 tgtgtgtgtg cgtggggagc gccgcgtgcg gccgcgctg cccggcggt gtgagcgctg 960
 cgggcgcggc gcggggcttt gtgcgctccg cgtgtgcgag aggggagcgc ggcggggggc 1020
 ggtgccccgc ggtgcggggg ggctgcgagg ggaacaaagg ctgctgctg ggtgtgtgct 1080
 tgggggggtg agcagggggt gtgggcgcgg cggctcgggt gtaaccccc cctgcacccc 1140
 cctccccgag ttgctgagca cggccccggt tcgggtgcgg ggctccgtgc ggggcgtggc 1200
 gcggggctcg ccgtccggg ccgggggggtgg cggcaggtgg ggggtgccgg cggggcgggg 1260
 ccgcctcggg ccggggaggg ctccgggggag ggcgcggcg gccccggagc gccggcggt 1320
 gtgcagggcg ggcgagccgc agccattgcc ttttatggta atcgtgcgag agggcgccagg 1380
 gacttccttt gtcccaaatc tggcggagcc gaaatctggg aggcgcggcc gcacccccct 1440
 tagcggcgcg gggcgaaagc gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt 1500
 cgtgctgcgc cgcgcgcggc tccccttctc cgggggtcgg cgggggtcgg ctccgggctg ccgcaggggg 1560
 acggctgcct tcggggggga cggggcgagg cggggtcgg cttctggcgt gtgacggcg 1620
 gctctagagc ctctgtaac ctctgtcatg ccttcttctt tttcctacag ctccctggga 1680
 acgtgctggt tgttgtgtg tctcatcatt ttggcaaga attcactcct cagggtgcagg 1740
 ctgcctatca gaagtggtg gctgggtgtg ccaatgccct ggctcacaaa taccactgag 1800
 atctttttcc ctctgccaaa aattatgggg acatcatgaa gcccttgag catctgactt 1860
 ctggctaata aaggaaattt attttcattg caatagtgtg ttggaatttt ttgtgtctct 1920
 cactcggaag gacatatggg agggcaaatc atttaaaaca tcagaatgag tatttgggtt 1980
 agagtttggc aacatatgcc cctgctgtc tgccatgaac aaaggtggct ataaagagg 2040
 catcagata tgaaacagcc cctgctgtc cactccttat tccatagaaa agccttgact 2100
 tgagggtaga ttttttttat attttgttt ttctcctct cctgactact cccagtcata 2220
 ttttccttac atgttttact agccagattt gacctgcagc ccaagcttgg cgtaatcatg 2280
 gctgtccctc ttctcttatg aagatccctc gaaattgtta tccgctcaca attccacaca acatacgagc 2340
 gtcatagctg tttcctgtgt cctgggggtg ctaatgagtg agctaactca cattaattgc 2400
 cggaagcata aagtgtaaaag tccagtcggg aaacctgtcg tgccagcggg tccgcatctc 2460
 gttgcgctca ctgccgctt ccagtcagg cccgccccca tcccgccct aactccgccc 2520
 aattagtcag caaccatagt cccgccccca ctatgttatt ttatttatgc agagggcgag 2580
 agttccgccc attctccgcc ccatggctga attccagaag gcttttttgg aggcctaggc 2640
 gccgcctcgg cctctgagct tttattgcag cttataatgg ttacaaataa agcaatagca 2700
 ttttgcaaaa agctaacttg tttattgcag cactgcattc tagttgtggt ttgtccaaac 2760
 tcacaaattt cacaaataaa gcattttttt gctctggatc gctgcattaa tgaatcgggc aacgcgcggg 2820
 tcatcaatgt atcttatcat gtctggatcc cgtctcctcg ctactgact cgctgcgctc 2880
 gagaggcggt ttgcgtattg ggcgctcttc cgtatcagc tcaactcaaag gcgtaatac ggttatccac 2940
 ggtcgttcgg ctgcggcgag cggatcagc gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa 3000
 agaatcaggg gataacgcag tggcggtttt ccataggctc aaacccgaca cgcccccctg acgagcata 3060
 ccgtaaaaag gccgcgttgc agaggtggcg agactataaa gataccaggc 3120
 caaaaaatcg cgctcaagtc tegtgcgctc tctgtttccg accctgcccgc ttaccggata 3180
 gtttccccct ggaagctccc tegtgcgctc

-37-

cctgtccgcc	tttctccctt	cgggaagcgt	ggcgctttct	caatgctcac	gctgtaggta	3240
tctcagttcg	gtgtagggtcg	tctcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	3300
gcccgaaccgc	tgcgctttat	cgggtaacta	tctgtcttgag	tccaaccccg	taagacacga	3360
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgagggt	atgtaggcgg	3420
tgctacagag	ttcttgaagt	gggtggcctaa	ctacggctac	actagaagga	cagtatttgg	3480
tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	3540
caaacaaacc	accgctggta	gcggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	3600
aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	3660
cgaaaactca	cgtaaaggga	ttttggtcat	gagattatca	aaaaggatct	tcacctagat	3720
ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	atataatgagt	aaacttggtc	3780
tgacagttac	caatgcttaa	tcagtgaggc	acctatctca	gcgatctgtc	tatttcgttc	3840
atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	3900
tggccccagt	gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atttatcagc	3960
aataaaccag	ccagccggaa	gggcccagcg	cagaagtggg	cctgcaactt	tatccgcctc	4020
catccagttc	attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatatggt	4080
gcgcaacgtt	gttgccattg	ctacaggcat	cgtgggtgtca	cgctcgtcgt	ttggtatggc	4140
ttcattcagc	tccggttccc	aacgatcaag	gcgagctaca	tgatccccc	tggtgtgcaa	4200
aaaagcgggt	agctccttcg	gtcctccgat	cggtgtcaga	agtaagttgg	ccgcagtggt	4260
atcactcatg	gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	4320
cttttctgtg	actgggtgag	actcaaccac	gtcattctga	gaatagtgtg	tgccggcacc	4380
gagttgctct	tgcccggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	4440
agtgtctatc	atgggaaaac	gttcttcggg	gcgaaaaactc	tcaaggatct	taccgctgtt	4500
gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	4560
caccagcggt	tctgggtgag	caaaaacagg	aaggcaaaat	gccgcaaaaa	aggggaataag	4620
ggcgacacgg	aaatgttgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	4680
tcagggttat	tgtctcatga	gcggtacat	atttgaatgt	atttagaaaa	ataaacaaat	4740
aggggttccg	cgcacatttc	cccgaaggt	gccacctg			4778

<210> 71

<211> 5510

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFP plasmid

<400> 71

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataatat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgccct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgtg	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccc	300
cctggcatta	tgcccagtag	atgaccttat	gggaactttcc	tacttggcag	tacatctacg	360
tattagtcac	cgtctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctccc	420
atctccccc	cctccccacc	cccaattttg	tattttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccagggcgg	gcggggcggg	gcgagggggc	540
ggggcgggcg	aggcgggag	gtgcggcggc	agccaatcag	agcggcgcg	tccgaaagtt	600
tcccttttat	gcgagggcgg	ggcggcggcg	gccctataaa	aagcgaagcg	cgcggcgggc	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgctccgc	gccgcctcgc	gccgcccgc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgctgaaaag	840
ccttaaaggg	ctccgggagg	gccctttgtg	cgggggggag	cggctcgggg	gggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtcgc	gccgcgcgtg	cccggcgggt	gtgagcgcgtg	960
cgggccccgc	ggcgggcttt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggcggggggc	1020
gggtccccgc	ggtgcggggg	ggctgcgagg	ggaacaaaagg	ctgcgtgcgg	gggtgtgtgcg	1080
tggggggggt	agcagggggg	gtgggcgcgg	cggctcgggt	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggccccggt	tgggtgcgg	ggctccgtgc	ggggcgtggc	1200
cgggggctcg	ccgtgcgggg	cgggggggtg	cggcaggtgg	gggtgcgggg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggtc	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgcagg	1380
gacttctctt	gtcccaaatc	tggcggagcc	gaaatctggg	aggcgcgcgc	gcacccctc	1440
tagcggggcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acgggtgcct	tccgggggga	cggggcaggg	cgggggttcg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatt	ccttctctct	tttctacag	ctcctgggca	1680
acgtgctggg	tgttgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atgggtgagca	1740
agggcgagga	gctgttcacc	gggggtggtg	ccatcctggt	cgagctggac	ggcgacgtaa	1800

-38-

acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accgggcaagc	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgaccta	cggcgtgcag	tgcttcagcc	gctaaccgga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgac	gcccaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tgggccgaaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcgagcg	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcgac	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatggtc	ctgctggagt	2400
tcgtgaccgc	cgccgggac	actctcggca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaaggtgg	tggtctgggt	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccccctg	2580
agcatctgac	ttctggctaa	ttaaggaaat	ttattttcat	tgcaatagtg	tggtggaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcatttataa	catcagaatg	2700
agtattttgt	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaaggtgg	2760
ctataaagag	gtgtcatagta	tatgaaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atatttttgt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttcctt	acatgtttta	ctagccagat	ttttcctcct	ctcctgacta	2940
ctccccagtc	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gccccagctt	3000
ggcgtaatca	tggtcctagt	tggttccctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gcccgaagca	taaagtgtaa	agcctggggg	gcctaagtga	tgagctaact	3120
cacattaatg	gcgttgcgtc	cactgccccg	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcatc	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgcgc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttatttat	3300
gcagagggcg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggagggcctag	gcttttgcaa	aaagctaact	tggttattgc	agcttataat	ggttacaaat	3420
aaagcaaatg	catcacaatg	ttcacaaata	aagcattttt	ttcactgcat	tctagttgtg	3480
gtttgtccaa	actcatcaat	gtatcttatc	atgtctggat	ccgctgcatt	aatgaatcgg	3540
ccaacgcgcg	gggagagggc	gtttgctgat	tgggcgctct	tccgcttcct	cgctcactga	3600
ctcgctgcgc	tcggtcgttc	ggctgcggcg	agcggtatca	gctcactcaa	aggcggtaat	3660
acgggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3720
aaaggccagg	aaccgtaaaa	aggccgcgct	gctggcgctt	ttccataggc	tccgcccccc	3780
tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	cgaaccgga	caggactata	3840
aagataaccg	cggtttcccc	ctcgtgcgc	cctcgtgcgc	tctcctgttc	cgaccctgcc	3900
gcttaccgga	tacctgtccg	cctttctccc	ttcgggaagc	gtggcgcttt	ctcaatgctc	3960
acgctgtagg	tatctcagtt	cggtgtaggt	cgttctgctc	aagctgggct	gtgtgcacga	4020
accccccggt	gaccccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaaccc	4080
ggtaagacac	gacttatcgc	cactggcagc	agccactggg	aacaggatta	gcagagcgag	4140
gtatgtaggc	ggtgctacag	agttcttgaa	gtgggtggct	aactacggct	acactagaag	4200
gacagtattt	ggtatctgcg	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4260
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgggtgg	ttttttgttt	gcaagcagca	4320
gattacgcgc	agaaaaaaag	gatctcaaga	agatcccttg	atcttttcta	cggggtctga	4380
cgctcagttg	aacgaaaact	cacgttaagg	gatttttggt	atgagattat	caaaaaggat	4440
cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	tcaatctaaa	gtatatatga	4500
gtaaaacttg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4560
tctatttcgt	tcatccatag	ttgcctgact	ccccgtcgtg	tagataacta	cgatacggga	4620
gggcttacc	tctggcccca	gtgctgcaat	gatacccgga	gacccacgct	caccggctcc	4680
agattttatc	gcaataaacc	agccagccgg	aaggggccgag	cgcagaagtg	gtcctgcaac	4740
tttatccgac	tccatccagt	ctattaattg	ttgccgggaa	gctagagtaa	gtagttcgcc	4800
agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	atcgtgggtg	cacgctcgte	4860
gtttgggtatg	gcttcattca	gctccggttc	ccaacgatca	aggcgagtta	catgatcccc	4920
catgtttgtgc	aaaaaagcgg	ttagctcctt	cggtcctccg	atcgttgtca	gaagtaagtt	4980
ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgcc	5040
atccgtaaga	tgcttttctg	tgactgggtg	gtactcaacc	aagtcattct	gagaatagtg	5100
tatgcccga	ccgagttgct	cttgcccggc	gtcaataacg	gataataccg	cgccacatag	5160
cagaacttta	aaagtgtctc	tcattggaaa	acgttcttcg	gggcgaaaac	tctcaaggat	5220
cttaccgctg	ttgagatcca	gttcgatgta	acccactcgt	gcacccaact	gatcttcagc	5280
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaaca	ggaaggcaca	atgccgcaaa	5340
aaaggggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttctctt	ttcaatatta	5400
ttgaagcatt	tatcaggggt	attgtctcat	gagcggatac	atatttgaat	gtatttagaa	5460
aaataaacaa	ataggggttc	cgcgcacatt	tccccgaaaa	gtgccacctg		5510

<210> 72

<211> 282

<212> DNA

<213> Artificial Sequence

-39-

<220>

<223> attp

<400> 72

ccttgcgcta	atgctctgtt	acagggtcact	aataccatct	aagtagttga	ttcatagtga	60
ctgcatatgt	tgtgttttac	agtattatgt	agtctgtttt	ttatgcaaaa	tctaatttaa	120
tatatggata	tttatatcat	tttacgtttc	tcgttcagct	tttttatact	aagttggcat	180
tataaaaaag	cattgcttat	caatttggtg	caacgaacag	gtcactatca	gtcaaaaataa	240
aatcattatt	tgatttcaat	tttgtccac	tccctgcctc	tg		282

<210> 73

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 73

ggccccgtaa	tcgagaagaa	20
------------	------------	----

<210> 74

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 74

ggtttaaagt	gcgctcctcc	aagaacgtca	tc	32
------------	------------	------------	----	----

<210> 75

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 75

agatctagag	ccgccgctac	aggaacaggt	ggaggcgggc	40
------------	------------	------------	------------	----

<210> 76

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer 5PacSV40

<400> 76

ctgttaatta	actgtggaat	gtgtgtcagt	taggggtg	37
------------	------------	------------	----------	----

<210> 77

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer Antisense Zeo

<400> 77

tgaacagggt	cacgtcgtcc	20
------------	------------	----

<210> 78

<211> 24

-40-

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5' HETS

<400> 78
gggccgaaac gatctcaacc tatt 24

<210> 79
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3' HETS

<400> 79
cgcagcggcc ctcctactc 19

<210> 80
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5BSD

<400> 80
accatgaaaa catttaacat ttctcaaca 29

<210> 81
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer SV40polyA

<400> 81
tttatttgtg aaatttgtga tgctattgc 29

<210> 82
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3BSP

<400> 82
ttaatttcgg gtatatattga gtgga 25

<210> 83
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer EPO5XBA

<400> 83
tatctagaat ggggggtgcac gaatgtcctg cc 32

<210> 84
<211> 32
<212> DNA

-41-

<213> Artificial Sequence

<220>

<223> Primer EPO3SBI

<400> 84

tacgtacgtc atctgtcccc tgtcctgcag gc

32

<210> 85

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEPO3BSI

<400> 85

cgtacgtcat ctgtcccctg tcctgca

27

<210> 86

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEPO5XBA

<400> 86

tctagaatgg ggggtgcacgg tgagtact

28

<210> 87

<211> 4862

<212> DNA

<213> Artificial Sequence

<220>

<223> pD2eGFP-1N plasmid from Clontech

<400> 87

tagttattaa	tagtaatcaa	ttacgggggtc	attagttcat	agcccatata	tggagttccg	60
cggtacataa	cttacggtaa	atggcccgc	tggctgaccg	cccaacgacc	cccggccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacggt	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtaacgcc	cctattgacg	tcaatgacgg	taaattggccc	gcctggcatt	atgccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgt	acaactccgc	ccatttgacg	caaattgggcg	gtaggcgtgt	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacggtagc	gcccggcccg	660
gatccaccgg	tcgccaccat	ggtgagcaag	ggcgaggagc	tggtcaccgg	ggtgggtgcc	720
atcctgggtc	agctggacgg	cgacgtaaac	ggccacaagt	tcagcgtgtc	cggcgagggc	780
gagggcgatg	ccacctacgg	caagctgacc	ctgaagttca	tctgcaccac	cggcaagctg	840
cccgtgccct	ggcccaccct	cgtgaccacc	ctgacctacg	gcgtgcagtg	cttcagccgc	900
taccccggacc	acatgaagca	gcacgacttc	ttcaagtccg	ccatgccga	aggctacgtc	960
caggagcgca	ccatcttctt	caaggacgac	ggcaactaca	agacccgcgc	cgaggtgaag	1020
ttcgaggggc	acaccctggt	gaaccgcac	gagctgaagg	gcatcgactt	caaggaggac	1080
ggcaacatcc	tggggcacaa	gctggagtac	aactacaaca	gccacaacgt	ctatatcatg	1140
gccgacaagc	agaagaacgg	catcaaggtg	aacttcaaga	tccgccacaa	catcgaggac	1200
ggcagcgtgc	agctcgccga	ccactaccag	cagaacaccc	ccatcgccga	cggccccgtg	1260
ctgctgcccc	acaaccacta	cctgagcacc	cagtccgccc	tgagcaaaga	ccccaacgag	1320
aagcgcgcatc	acttggtcct	gctggagttc	gtgaccgcgc	cggggtcac	tctcggcattg	1380
gacgagctgt	acaagaagct	tagccatggc	ttcccgcggg	aggtggagga	gcaggatgat	1440
ggcagcgtgc	ccatgtcttg	tgcccaggag	agcgggatgg	accgtcacc	tgacgcctgt	1500
gcttctgcta	ggatcaatgt	gtagatgcgc	ggccgcgact	ctagatcata	atcagccata	1560
ccacatttgt	agaggtttta	cttgctttta	aaaaacctcc	acacctcccc	ctgaacctga	1620
aacataaaat	gaatgcaatt	gttggtgtta	acttggttat	tgcagcttat	aatgggttaca	1680

-42-

```

aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 1740
gtgggtttgtc caaactcatc aatgtatctt aaggcgtaaa ttgtaagcgt taatattttg 1800
ttaaaattcgc cgttaaatatt ttgttaaatac agctcatttt ttaaccaata ggccgaaatc 1860
ggcaaaatccc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt 1920
tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaaggcg aaaaaccgtc 1980
tatcagggcg atggcccact acgtgaacca tcacctaat caagtttttt ggggtcgagg 2040
tgccgtaaag cactaatcgc gaaccctaaa gggagccccc gatttagagc ttgacgggga 2100
aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg 2160
ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgcgc 2220
ctacagggcg cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc tatttgttta 2280
tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg ataatgtctt 2340
caataatatt gaaaaagga gagtcctgag gcggaagaa ccagctgtgg aatgtgtgtc 2400
agttagggtg tggaaagtcc ccagggtccc caggcaggcag aagtatgcaa agcatgcatc 2460
tcaattagtc agcaaccagg tgtggaaagt cccagggtc cccagcaggc agaagtatgc 2520
aaagcatgca tctcaattag tcagcaacca tagtcccgcc cctaactccg cccatcccgc 2580
ccctaactcc gccagttcc gccatttct agtccccatgg ctgactaatt ttttttattt 2640
atgcagagggc caggcccgcc tcggcctctg agctattcca gaagttagtga ggaggccttt 2700
ttggaggcct aggtcttttg aaagatcgat caagagacag gatgaggatc gtttcgcgat 2760
attgaacaag atggattgca cgcaggttct ccggccgctt ggttgagag gctattcggc 2820
tatgactggg cacaacagac aatcggctgc tctgatgccg ccgtgttccg gctgtcagcg 2880
caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa tgaactgcaa 2940
gacgaggcag cgggctatc gtggctggcc acgacgggcg ttccttgccg agctgtgtc 3000
gacgttgtca ctgaagcggg aagggaactgg ctgctattgg gcgaagtgcc ggggcaggat 3060
ctcctgtcat cggttgatcc tctgcccag aaagtatcca tcatggctga tgcaatgcgg 3120
cggctgcata cggttgatcc ggctacctgc ccattcgacc accaagcgaa acatcgcatc 3180
gagcgagcac gtactcggat ggaagccggg cttgtcgatc aggatgatct ggacgaagag 3240
catcaggggc tcgcgccagc cgaactgttc cccaggctca aggcgagcat gccgcagcgc 3300
gaggatctcg tctgaccca tggcgatgcc tgcttgccga atatcatggg ggaaaatggc 3360
cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata 3420
gcgttggtca ccgtgatat tgctgaagag attgggctga ccttctatcg ccgcttcctc 3480
gtgctttacg gtatcgccgc tcccgatctg cagcgcctcg ccttctatcg ccttcttgac 3540
gagttcttct gagcgggact ctgggggttcg aaatgaccga ccaagcgacg cccaacctgc 3600
catcacgaga tttcgattcc accgccgcct tctatgaaag gttgggcttc ggaatcgttt 3660
tccgggacgc cggctggatg atcctccagc gcggggatct catgctggag ttcttcgccc 3720
accctagggg gaggctaact gaaacacgga agggagacaat accggaagga acccgcgcta 3780
tgacggcaat aaaaagacag aataaaacgc accggtgttg gtctgttgtt cataaacgcg 3840
gggttcggtc cagggtcgg cactctgtcg ataccccacc gagaccccat tggggccaat 3900
acgcccgcgt tcttctctt tccccacccc acccccgaag ttcggtgaa ggcccagggc 3960
tcgcagccaa cgtcggggcg gcaggccctg ccatagcctc aggttactca tatatacttt 4020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 4080
atctcatgac caaaatccct taacgtgagt ttctgtcca ctgagcgtca gaccccgtag 4140
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa 4200
caaaaaaac accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 4260
ttcgaaggt aactggcttc agcacacaaa tactgtcctt ctagttagc 4320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa 4380
tctgttacc agtggctgct gccagtggcg tgagatacct tcttaccggg ttggactcaa 4440
gacgatagtt accggataag ggcagcgggt cgggctgaac ggggggttcg tgcacacagc 4500
ccagcttgga gcgaacgacc tacaccgaac acaggtatcc ggtaagcggc ctatgagaaa 4560
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtatctttat agggctcgaa 4620
caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 4680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctctcagggg gggcgagcc 4740
tatggaaaaa cgccagcaac ggggcctttt tacggttcct ggcttttgc tggccttttg 4800
ctcacatggt ctttctctgc ttatcccctg attctgtgga taaccgtatt accgccatgc 4860
at

```

<210> 88

<211> 5192

<212> DNA

<213> Artificial Sequence

<220>

<223> pIRESpuo2 plasmid from Clontech

<400> 88

```

gacggatcgg gagatctccc gatcccctat ggtcgactct cagtacaatc tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

```


-43-

ttaggggttag	gcgtttttg	ctgcttcg	atgtacggg	cagatatac	cgttgacatt	240
gattattgac	tagttattaa	tagtaatac	ttacggggg	attagttcat	agccccata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgc	tggctgaccg	cccaacgacc	360
cccgcccat	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggg	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaattggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaattggcg	780
gtaggcggtg	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttggtacc	900
gagctcggat	cgatatctgc	ggcctagcta	gcgcttaagg	cctgttaacc	ggtcgtaagt	960
ctccggattc	gaattcggat	ccgcggccgc	atagataact	gatccagttg	gctggaatta	1020
attcgctgtc	tgcgagggcc	agctgttggg	gtgagtactc	cctctcaaaa	gcgggcatga	1080
cttctgcgct	aagattgtca	gtttccaaaa	acgaggagga	tttgatattc	acctggcccg	1140
cggtgatgcc	tttaggggtg	gccgcgtcca	tctggtcaga	aaagacaatc	ttttgttgtt	1200
caagcttgag	gtgtggcagg	cttgagatct	ggccatacac	ttgagtgaca	atgacatcca	1260
ctttgccttt	ctctccacag	gtgtccactc	ccaggtccaa	ctgcaggctg	agcatgcatc	1320
tagggcgggc	aattccgccc	ctctccctcc	ccccccctta	acgttactgg	ccgaagccgc	1380
ttggaataag	gcccgtgtgc	gtttgtctat	atgtgatttt	ccaccatatt	gccgtctttt	1440
ggcaatgtga	gggcccggaa	acctggccct	gtcttcttga	cgagcattcc	taggggtctt	1500
tccctctctg	ccaaaggaat	gcaaggtctg	ttgaatgtcg	tgaaggagc	agttcctctg	1560
gaagcttctt	gaagacaaac	aacgtctgta	gcgacccttt	gcaggcagcg	gaacccccca	1620
cctggcgaca	ggtgcctctg	cggccaaaag	ccacgtgtat	aagatacacc	tgcaaaggcg	1680
gcacaacccc	agtgcacagt	tgtgagttgg	atagttgtgg	aaagagtcaa	atggctctcc	1740
tcaagcgtat	tcaacaaggg	gctgaaggat	gcccagaagg	tacccattg	tatgggatct	1800
gatctggggc	ctcgggtgcac	atgctttaca	tgtgtttagt	cgaggttaaa	aaaacgtcta	1860
ggccccccga	accacgggga	cgtggttttc	ctttgaaaaa	cacgatgata	agcttgccac	1920
aacccacaag	gagcagacct	tccatgaccg	agtacaagcc	cacgggtgoc	ctcgccaccc	1980
gcgcacatgg	ccccggggc	gtacgcaccc	tcgcgcgcgc	gttcgcccgc	taccccgcca	2040
cgcgccacac	cgtcgacccg	gaccgccaca	tcgagcgggt	caccgagctg	caagaactct	2100
tcctcacgcg	cgtcgggctc	gacatcggca	aggtgtgggt	cgcgagcagc	ggcgccggcg	2160
tggcggtctg	gaccacgcgc	gagagcgtcg	aagcgggggc	ggtgttcgcc	gagatcgccc	2220
cgcgcatggc	cgagttaggc	ggttcccgcc	tggccgcgca	gcaacagatg	gaaggcctcc	2280
tggcgccgca	ccggcccaaag	gagcccgcgt	ggttcctggc	caccgtcggc	gtctcgcccg	2340
accaccaggg	caagggtctg	ggcagcggcg	tcgtgctccc	cggagtggag	gcgcccgagc	2400
gcgcgggggt	gcccgccttc	ctggagacct	ccgcgcctcc	caacctcccc	ttctacgagc	2460
ggctcggctt	caccgtcacc	gcccagctcg	agtgcgccga	ggaccgcgcg	acctggtgca	2520
tgaccgcgaa	gcccgggtgc	tgacgcgccg	cccaocgccc	gcagcgcccc	accgaaagga	2580
gcgcacgacc	ccatggctcc	gaccgaagcc	gacccgggcg	gccccggcga	ccccgcaccc	2640
gcccccgagg	cccaccgact	ctagagctcg	ctgatcagcc	tcgactgtgc	cttctagtgt	2700
ccagccatct	gttgtttgcc	cctccccctg	gccttccttg	acctgggaag	gtgccactcc	2760
cactgtcctt	tcctaataaa	atgaggaat	tgcatcgcac	tgtctgagta	ggtgtcattc	2820
tattctgggg	ggtgggggtg	ggcaggacag	caagggggag	gattgggaag	acaatagcac	2880
gcatgctggg	gctcgtgggt	gctctatggc	ttctgaggcg	gaaagaacca	gctggggctc	2940
gagtgcattc	tagttgtggg	ttgtccaaac	tcataaatgt	atcttatcat	gtctgtatac	3000
cgtcgacctc	tagctagagc	ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	3060
gttatccgct	cacaattcca	cacaacatac	gagccgggaag	cataaagtgt	aaagcctggg	3120
gtgcctaata	agtgcgctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	gctttccagt	3180
cgggaaaacct	gtcgtgccag	ctgcattaat	gaatcgccca	acgcgcgggg	agaggcggtt	3240
tgcgtattgg	gcgctcttcc	gcttctctcg	ctactgactc	gctgcgctcg	gtcgttcggc	3300
tgcggcgagc	ggtatcagct	cactcaaagg	cggtaatagc	gttatccaca	gaatcagggg	3360
ataacgcagg	aaagaacatg	tgagcaaaaag	gccagcaaaa	ggccaggaac	cgtaaaaagg	3420
cccggttgct	ggcggttttt	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	3480
gctcaagtca	gaggtggcga	aacccgacag	gactataaaag	ataccaggcg	tttccccctg	3540
gaagctccct	cgtgcgctct	cctgttccga	ccctgcccgt	taccggatac	ctgtccgcct	3600
ttctcccttc	gggaagcgtg	gcgctttctc	aatgctcacg	ctgtagggtat	ctcagttcgg	3660
tgtaggctcg	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	cccgaacgct	3720
gcgccttata	cggtaactat	cgtcttgagt	ccaacccggg	aagacacgac	ttatcgccac	3780
tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	3840
tcttgaagtg	gtggccctaac	tacggctaca	ctagaaggac	agtatttggt	atctgcgctc	3900
tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	3960
ccgctggtag	cggtgggttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	4020
ctcaagaaga	tcctttgatc	ttttctacgg	ggcttgacgc	tcagtggaac	gaaaactcac	4080
gttaagggat	tttggctatg	agattatcaa	aaaggatctt	cacctagatc	cttttaatt	4140
aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtc	gacagttacc	4200

-44-

```

aatgcttaat cagtgaaggca cctatctcag cgatctgtct atttcgttca tccatagttg 4260
cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg 4320
ctgcaatgat accgcgagac ccacgctcac cggtccaga tttatcagca ataaaccagc 4380
cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc atccagtcta 4440
ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg 4500
ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct tcattcagct 4560
ccggttcccc cagatcaagg cgagttacat gatcccccac gttgtgcaaa aaagcgggta 4620
gctccttcgg tcctccgacg gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg 4680
ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga 4740
ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccc agttgctctt 4800
gcccgcgctc aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca 4860
ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt 4920
cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt 4980
ctgggtgagc aaaaaacgga aggcataaat gcgcaaaaaa gggaataagg gcgaacgga 5040
aatgttgaat actcatactc ttcttttttc aatattattg aagcatttat cagggttatt 5100
gtctcatgag cggatacata tttgaatgta tttgaaaaaa taaacaaata ggggttccgc 5160
gcacatttcc ccgaaaagtg ccacctgacg tc 5192

```

<210> 89

<211> 11182

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg1 Plasmid

<400> 89

```

catgccaaac acagggttcc cctcgggacg aaagtacttt gatccaaccg ctccgctgct 60
atagtgcagt cggcttctga cgttcagtgc agccgtcttc tgaaaacgac atgtcgcaca 120
agtcctaagt tacgcgacag gctgcgcccc tgcccttttc ctggcgcttt ctgtcgcgt 180
gttttagtgc cataaagtag aatacttgcg actagaaccg gagacattac gccatgaaca 240
agagcgccgc cgctggcctg ctgggctatg ccgcgctcag caccgacgac caggacttga 300
ccaaccaacg ggccgaactg cacgcggccg gctgcaccaa gctgttttcc gagaagatca 360
ccggcaccag gcgcgaccgc ccggagctgg ccaggatgct tgaccaccta cgccctggcg 420
acgcttgtag agtgaccagg ctgaccgccg tggcccgtag caccgcgac ctactggaca 480
ttgccgagcg catccaggag gccggcgccg gcctgcgtag cctggcagag ccgtgggccc 540
acaccaccac gccggccggc cgcattggtg tgaccgtgtt cgccggcatt gccagttcgc 600
agcgttccct aatcatcgac cgaccccgga ggcggcgaga ggcggccaag gcccgaggcg 660
tgaagtttgg ccccgccctt accctcaccg cggcacagat cgcgacgccc cgcgagctga 720
tcgaccagga aggcgcgacc gtgaaagagg cggctgcact gcttggcggt gctcgtcga 780
ccctgtaccg cgcacttgag cgcagcgagg aagtgcgccc caccgaggcc aggcggcgcg 840
gtgcttcccg ttgaccgagg ttgaccgagg cgacgcccct ggccggcgcc gagaatgaac 900
gccaagagga acaagcatga aaccgcacca ggacggccag gacgaaccgt ttttcattac 960
cgaagagatc gaggcggaga tgatcgcgcc cgggtacgtg ttcgagccgc ccgcgcacgt 1020
ctcaaccgtg cggctgcagt aaatcctggc cggtttgtct gatgccaagc tggcggcctg 1080
gccggccagc ttggccgctg aagaaaccga gcgcccgcgt ctaaaaaggt gatgtgtatt 1140
tgagtaaaac agcttgcgct atcggttcgc tgcgtatatg atgcgatgag taaataaaca 1200
aatacgcaag gggaacgcac gaaggttatc gctgtactta accagaaagg cgggtcaggg 1260
aagacgacca tcgcaaccca tctagccccg gccctgcaac tcgccggggc cgtgttctcg 1320
ttagtcgatt ccgatcccca gggcagtgcc cgcgattggg cggccgtgcg ggaagatcaa 1380
ccgctaaccg ttgtcggcat cgaccgcccc acgattgacc gcgacgtgaa ggccatcggc 1440
cggcgcgact tcgtagtgat cgacggagcg ccccaggcgg cggacttggc tgtgtccgcg 1500
atcaaggcag ccgacttcgt gctgattccg gtgcagccaa gcccttacga catatggggc 1560
accgcccagg tggtggagct ggttaagcag cgcattgagg tcacggatgg aaggctacaa 1620
gcgctggccg tcgtgtcgcg ggcgatcaaa ggcacgcgca tcggcggtga ggttgcgag 1680
gcgctggccg ggtacgagct gccattctt gagtccgta tcacgcagcg cgtgagctac 1740
ccaggcactg ccgcgcggcg cacaaccggt cttgaatcag aacccgaggg cgacgctgcc 1800
cgcgaggtcc aggcgctggc cgtgaaatt aaatcaaac tcatttgagt taatgaggta 1860
aagagaaaat gagcaaaagc acaaacacgc taagtgccgg ccgtccgagc gcacgcagca 1920
gcaaggctgc aacggttggcc agcctggcag acacgccagc catgaagcgg gtacgcgcaa 1980
agttgcggcg ggaggatcac accaagctga agatgtacgc ggtacgcgca gtaaatgagc 2040
ttaccagagt gctatctgaa tacatcgcgg agctaccaga gtaaatgagc acaaccaggc 2100
atgagtagat gaatttttagc ggctaaagga ggcggcatgg aaaatcaaga acaccaggc 2160
accgacggcg tggaatgccc catgtgtgga ggaacgggag gttggccagg cgtgaagcggc 2220
tggtgtgtct gccggccctg caatggcact ggaaccccca agcccgagga atcgcgctga 2280
cggtcgcaaa ccacccggcc cggtacaaat cggcgcgcg gaggcagaag caccggccgg 2340
gaagttgaag gccgcgcagg ccgcccagcg gcaacgcacg gatgcagaa 2400

```

-45-

tgaatcgtgg	caagcgcccg	ctgatcgaat	ccgcaaagaa	tcccggcaac	cgccggcagc	2460
cgggtgcgcg	tgcattagga	agccgcgcaa	ggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tcgcagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagaggtt	tccgcagggc	cggccggcat	ggccagtgtg	tgggattacg	acctggtact	2700
gatggcgggt	tcccattctaa	ccgaatccat	gaaccgatac	cgggaaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagtctctgc	ggcgagccga	2820
tggcggaaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgagggtga	2940
agccttgatt	agccgtaca	agatcgtaaa	gagcgaaacc	ggcgggcccg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggaagctgt	3060
gacggttcac	cccgtattct	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgcct	3120
ggcacgccgc	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagag	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atgtgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	caaatgtccc	tagcagggga	aaaaggctcg	aaaggctctc	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaccctg	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgctctaa	actctttaaa	acttattaaa	actcttaaaa	cccgctgggc	3600
ctgtgcataa	ctgtctggcc	agcgacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgttgcgc	tccctacgcc	ccgcgccttc	gcgtcggcct	atcgccggccg	ctggcccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accaggcgcc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cgccgccac	atcaaggcac	cctgcctcgc	gcgtttcggg	gatgacgggt	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagccca	3960
tgaccagctc	acgtagcgat	agccgagtg	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcggtg	tgaataacog	cacagatgcg	taaggagaaa	4080
ataccgcatc	aggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggtcggttcg	4140
gctgcggcga	ggcggtatcag	ctcactcaaa	ggcggtataa	cggttatcca	cagaatcagg	4200
ggataacgca	ggaagaaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtt	ctggcggttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaaccgcag	aggactataa	agataccagg	cgtttccccc	4380
tgaagctcc	ctcgctgcgt	ctcctgttcc	gaccctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtagggt	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgcttc	agcccgaccg	4560
ctgcgcctta	tcggtaact	atcgcttga	gtccaacccg	gtaagacacg	acttatcgcc	4620
actggcagca	ggcactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaaaaaac	4800
caccgctggg	agcgggtggt	tttttggttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	4920
acgttaaggg	attttggcta	tgcattctag	gtactaaaaa	aattcatcca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgagaca	5040
ctgttcttcc	ccgatatcct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgcccct	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgcggtggga	aaagacaagt	tcctcttcgg	gcttttcogt	5220
ctttaaaaaa	tcatacagct	cgcgcggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tcggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaagagacc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggct	ttgttcatct	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcattgcggt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggt	gtccctttat	accggctgtc	cgtcattttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	tattccttcc	gtatctttta	cgcagcggta	5700
tttttcgata	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttccct	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcaat	ctaaaacctt	aaataaccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	togttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6060
ctgcgcgctt	tacaacggct	ctcccgctga	cgcgctcccg	gactgatggg	ctgcctgtat	6120
cgagtgggtga	ttttgtgccc	agctgccggg	cggggagctg	ttggctgggt	ggtggcagga	6180
tattttgtgg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taattgtact	gaattaattc	gaattaatc	gggggatctg	gatttttagta	ctgggatttg	6300
gttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acatactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420

-46-

ggaactactc	acacattatt	atggagaaac	tgcagtcaaa	tctcgggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgcctcgtgc	6600
atgcgcacgc	tccgggtcgtt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcocgt	ccgctggtgg	6720
gggggggaga	cgtacacggt	cgactcggcc	gtccagtcgt	aggcgttgcg	tgcctccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccaggagtag	6840
cgctcccgcga	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgctg	ctcggtagcg	6900
aagttgaccg	tgctttgtctc	gatgtagtg	ttgacgatgg	tgacagaccg	cggcatgtcc	6960
gcctcggtag	cacggcggat	gtcggccggg	cgctcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatat	agaggaagg	cttgccgaagg	atagtgggat	tgtgctgcac	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaacgt	cttcttttct	7200
cacgatgctc	ctcgtgggtg	ggggcccatc	tttgggacca	ctgctgggac	aggcatcttg	7260
aacgatagcc	tttcccttat	cgcaatgatg	gcattttgtag	gtgccacctt	ccttttctac	7320
tgctcctttt	atgaaagtgc	agatagctgg	gcaatggaa	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggtcttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtg	cgtgctccac	catgttatca	catcaatcca	cttgccttga	7500
agacgtggtt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtaggtg	ccactctcct	tttctactgt	ccttttgatg	aagtgcacga	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctag	actgtatctt	tgatattott	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcagag	tgctctagcc	aatacgcaca	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgcacg	gtttcccgac	tggaaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tcgggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agaggtcgga	cggatatacag	acatgataag	atacatgtat	8100
gagtttgagc	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatattg	tgaaatttgt	8160
gatgtctatt	ctttatattgt	aaccattata	agctgcaata	aacaagttgg	gggtgggcga	8220
gaattgcagc	atgagatccc	cgcgtcggag	agctatccag	ccggcgtccc	ggaaaacgat	8280
tcggaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgctcc	tcggccaacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	ggctgctcgc	cgatctcggt	catggccggc	ccggaggcgt	cccgaagtt	8460
cgtggcacag	actccgac	actcgcgta	cagctcgtcc	aggccgcgca	cccacacctc	8520
ggccagggtg	ttgtccggca	ccacctggtc	ctggaccgcg	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acaccggcga	agtcgtcctc	cacgaagtcc	cgggagaacc	cgagccggtc	8640
ggtcgcagac	tcgacgcgtc	cggcgacgtc	agcaccggag	agcaccggaa	cggcactggt	8700
caacttgccc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaaaggg	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgccagac	tatctgtcac	ttcatcaaaa	cgcacagtaga	aaaggaaggt	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgccctctgcc	gacagtgtgc	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaaagca	agtggattga	ttgtgataca	tgggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggctat	tgagactttt	caacaaaggg	9180
taatatcggg	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagttagaaa	ggaagggtgg	acctacaaat	gcatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgcgac	agtggtccca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttctc	ctatataagg	9480
aagttcattt	cattttggaga	ggacacgctg	aaatcaccag	tctctctcta	caaactctatc	9540
tctctcgagc	tttcgcgagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gttcgatgt	aggagggcgt	ggatatgtcc	9720
tgccgggtaaa	tagctgcgac	gatggtttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcgccgcg	gctcccgatt	ccggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgtg	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgctgt	tctacaaccg	gtcgcggagg	ctatggatgc	gatcgctgcg	gccgatctta	9960
gccagacgag	cgggttcggc	ccattcggac	cgcaaggaa	cggccaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggagc	10080
acaccgtcag	tgcgctccgtc	gcgcaggctc	tcgatagact	gatgctttgg	gccgaggaca	10140
gccccgaagt	ccggcacctc	gtgcacggcg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcgggtc	attgactgga	gcgaggcgat	gttcggggat	tcaccaatagc	10260
aggtcgccaa	catcttcttc	tggaggccgt	ggttggtctg	tatggagcag	cagacgcgct	10320
acttcagcgc	gaggttcgag	gatcgccacg	actccgggcg	tatatgctcc	tatatgctcc	10380
gcattggtct	tgaccaactc	tatcagagct	tgggtgacgg	caatttcgat	gatgcagctt	10440

-47-

gggcgaggg	tcgatgagac	gcaatcggtc	gatccggagc	cgggactgtc	gggagtagac	10500
aaatgcggcg	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgcgata	10560
gtggaaaccg	acgccccagc	actcggtcga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	accaaatacc	10800
agtactaaaa	tccagatccc	ccgaattaat	tggcggttaa	ttcagatcaa	gcttggcact	10860
ggccgctggt	ttacaacgct	gtgactggga	aaaccctggc	gttaccacaac	ttaatcgctc	10920
tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	10980
ttcccaacag	ttgcgcagcc	tgaatggcga	atgctagagc	agcttgagct	tggatcagat	11040
tgtcgtttcc	cgcttccagt	ttaaactatc	agtgtttgac	aggatatatt	ggcgggtaaa	11100
cctaagagaa	aagagcggtt	attagaataa	cggatattta	aaagggcggt	aaaaggttta	11160
tccgttcgct	catttgtatg	tg				11182

<210> 90

<211> 8428

<212> DNA

<213> Artificial Sequence

<220>

<223> pCambia3300 Plasmid

<400> 90

catgccaacc	acagggttcc	cctcgggatc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggtcttctga	cggttcagtgc	agccgtcttc	tgaaaacgac	atgtcgacac	120
agtcctaagt	tacgcgacag	gctgcccggc	tgcccttttc	ctggcggttt	cttgtcgctg	180
gttttagtgc	cataaagtag	aatacttgcg	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggccgaactg	cacgcggccg	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tggcccgagc	cacccgcgac	ctactggaca	480
ttgccgagcg	catccaggag	gccggcgccg	gcctgctgag	cctggcagag	ccgtgggccc	540
acaccaccac	gccggcgccg	cgcatggtgt	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgcaccggga	gcgggcgcga	ggccgcgaag	gcccgaggcg	660
tgaagtttgg	ccccgccttc	accctcacc	cggcacagat	cgcgacgccc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaagagg	cggctgcact	gcttggcggt	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aagtgcagcc	caccgaggcc	aggcggcgcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	cgcagccctt	ggcgggccgc	gagaatgaac	900
gccaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcggaga	tgatcgccgg	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcgat	aaatcctggc	cggtttgtct	gatgccaaag	tggcggcctg	1080
gccggccagc	ttggccgctg	aagaaaccga	gcgcgcctgt	ctaaaaaggt	gatgtgtatt	1140
tgagtaaaac	agcttgccgtc	atgcggctgc	tgcttatatg	atgcgatgag	taataaataa	1200
aatacgaag	gggaacgcac	gaaggttatc	gctgtactta	accagaaagg	cgggtcaggc	1260
aagacgacca	tcgcaaccca	tctagcccg	gccctgcaac	tcgccggggc	cgatgttttg	1320
ttagtgcatt	ccgatcccca	gggcagtgcc	cgcgattggg	cgcccggtgc	ggaagatcaa	1380
ccgctaaccg	ttgtcgccat	cgaccgcccc	acgattgacc	gogacgtgaa	ggccatcggc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	ccccaggcgg	cggacttggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatggggc	1560
accgcgcacc	tggtggagct	ggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgccgag	1680
gcgctggccg	ggtagcagct	gcccattctt	gagtcctgta	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcccgcgg	cacaaccgtt	cttgaatcag	aacccgaggg	cgacgctgcc	1800
cgcgaggtcc	aggcgctggc	cgctgaaatt	aaatcaaaac	tcatttgagt	taatgaggtg	1860
aagagaaaat	gagcaaaaag	acaaacacgc	taagtgccgg	cgtcccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgcagc	catgaagcgg	gtcaacttct	1980
agttgccggc	ggaggatcac	accaagctga	agatgtacgc	ggtacgccaa	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgcg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagtagat	gaatttttag	ggctaaagga	ggcggtatgg	aaaatcaaga	acaaccaggc	2160
accgacgcgg	tggaaatgcc	catgtgtgga	ggaacggggc	gttggccagg	cgtaagcggc	2220
tgggttgtct	gcccggccctg	caatggcact	ggaaccccca	agcccgagga	atcgccgtga	2280
cggtcgcaaa	ccatccggcc	cggtagcaaat	cgccgcggcg	ctgggtgatg	acctgcttcc	2340
gaagttgaag	gccgcgcagg	ccgcccagcg	gcaacgcac	gaggcgagaag	cacgccccgg	2400
tgaatcgtag	caagcgcccg	ctgatcgaa	ccgcaagaa	tcccggcaac	cgccggcagc	2460
cgggtgcctg	tcgattagga	agccgcccga	cggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tcgcagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640

-48-

cgtagaggtt	tccgcagggc	cggccggcat	ggccagtggtg	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaagggg	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaaag	cagaaagacg	acctggtaga	aacctgcatt	cggtaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacgggat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	ggcgggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccgccatc	ggcggttttc	tctaccgcct	3120
ggcacgccgc	gccgcaggca	aggcagaagc	cagatggttg	ttcaagacga	tctacgacga	3180
cagtggcagc	gccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atttgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggc	caaattggcc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	atctgtggga	accggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tcctacgcgc	ccgcccgttc	gcgctcgccct	atcgccggcg	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cgccgcaccac	atcaaggcac	ctcgctcgc	gctgttcggg	gatgacgtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cggtgtgtgg	cgggtgtcgg	ggcgacagca	3960
tgaccagctc	acgttagcat	agcggagtg	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgacac	atatcggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcac	aggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggctcgttcg	4140
gctgcccga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcggttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataaccagg	cgtttccccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgccg	cttaccggat	acctgtccgc	4440
ggtgtaggct	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgcgtc	atctcagttc	4500
ctgcgcctta	tccggtaact	atcgtcttga	gtccaaccgc	gtaagacacg	agcccagccg	4560
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	acttatcgcc	4620
gttcttgaag	tggtagccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4680
tctgtgaag	ccagttacct	tcggaaaaaa	agttggtagc	tcttgatccg	gcaaaaaaac	4740
caccgctggg	agcggtaggt	ttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4800
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	4860
acgttaaggg	attttgggtca	tgcattctag	gtactaaaac	aattcoatcca	gtaaaaataa	4920
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	4980
ttctctttcc	ccgatctcct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccattt	5040
gtccgccttg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5100
gatgttgctg	tctcccaggt	cgccgtggga	aaagacaagt	tcctcttcgg	gcttttccgt	5160
ctttaaaaaa	tcatacagct	cgcgcgatc	tttaaatgga	gtgtotttct	cccagttttc	5220
gcaatccaca	tcggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5280
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5340
cgcatacage	tcgataatct	tttcagggct	ttgttcactc	tcatactctt	ccgagcaaa	5400
gacgccatcg	gcctcactca	tgagcagatt	gtccagcca	tcatgccgtt	caaagtgcag	5460
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5520
atcatagggt	gtccctttat	accggctgtc	cgtcattttt	aaatataggt	tttcattttc	5580
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggtg	5640
tttttcgac	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttccct	5700
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5760
aattcactgt	tccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaaagtg	5820
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5880
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	5940
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6000
tctgccgcct	tacaacggct	ctcccgtgta	gcgccgtccc	gactgatggg	ctgcctgtat	6060
cgagtggtga	ttttgtgccg	agctgccggg	cggggagctg	ttggctggct	ggtggcagga	6120
tatatgtggt	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggagctttt	6180
taattgtatg	aattaacgcc	gaattaatc	gggggatctg	gatttttagta	ctggattttg	6240
ggtttcttat	tttagaattt	tattgataga	agtattttac	aaatacaaat	acatactaa	6300
ggaaactact	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6360
ggacggggcg	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6420
ccgtgcttga	gtcagggcag	gctgaagtc	agctgccaga	aaaccacgtc	atgccagttc	6480
atgcgcacgc	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgcctcgtgc	6540
	tcgggtcggt	gggcagcccc	atgacagcga	ccacgctctt	gaagccctgt	6600
						6660

-49-

```

gcctccaggg acttcagcag gtgggtgtag agcgtggagc ccagtcocgt ccgctgggtgg 6720
cggggggaga cgtacacggg ggcagcccggt aggcgatgcc cgcgtccacgt aggcgatgcc 6780
gggcccgcgt gacggacgag gtcgtccgtc cactccctcgt gttcctgcgg ctcggtagcgg 6840
cgctcccgcga aagttgaccg tgcttgtctc ttgacgatgg tgcagaccgc cggcatgtcc 6900
gagatagatt tgttagagaga gactgggtgat ttcagcgtgt cctctccaaa tgaatgaac 7020
ttccttatat agaggaaggt cttgcgaagg atagtgggat tgtgcgtcat cccttacgtc 7140
agtggagata tcacatcaat ccacttgctt tgaagacgtg gttggaacgt cttctttttc 7200
cacgatgctc ctcggtgggtg ggggtccatc tttgggacca ctgtcggcag aggcattctt 7260
aacgatagcc tttcctttat cgcaatgatg gcatttgtag gtgccacctt ccttttctac 7320
tgtccttttg atgaagtgac agatagctgg gcaatggaat ccgaggaggt ttcccgatat 7380
taccctttgt tgaaaagtct caatagccct ttggtcttct gagactgtat ctttgatatt 7440
cttgagtag acgagagtgt cgtgctccac catgttatca catcaatcca cttgctttga 7500
agacgtgggt ggaacgtctt catcttgaac gatagcctt cctttatcgc aatgatggca 7560
gggaccactg tcggcagagg catcttgact tttctactgt ccttttgatg aagtgcaga 7620
tttgtaggtg ccaccttctt ccgatattac cctttgttga aaagtctcaa tagccctttg 7680
atggaatccg actgtatctt tgatattctt ggagtagacg agagtgtcgt gctccaccat 7740
gtctcttgag ggtggcaagc taatgcagct ggcacgacag agcgtcactca ggcgagcttg gccgattcat 7800
aatgtgagtt agctcactca gtttcccgac ttaggcacc cggcgtttac gcaagtgcag caacgcaatt 7860
atgtgtgtg taacacacag cggataacaa tttcacacag gaaacagcta tgcacatgat 7920
tacgaattcg agctcggtag agctcggtag tctagagtcg acctgcaggc atgcaagctt 7980
ggcactggcc gtcgtttttac aacgtcgtga ctgggaaaac cctggcggtt cccaaactta 8040
tcgccttgca gcacatcccc ctttcgccag ctggcgtaat agcgaagagg cccgcaccga 8100
tcgcccttcc caacagttgc gcagcctgaa tggcgaatgc tagagcagct ttagcttggg 8160
tcagattgtc gtttcccgcc ttcatgttaa actatcagtg tttgacagga tatattggcg 8220
ggtaaaccta agagaaaaga gcgtttatta gaataacgga tatttaaaag ggcgtgaaaa 8280
ggtttatccg ttcgtccatt tgtatgtg

```

<210> 91

<211> 3438

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBZeo Plasmid

<400> 91

```

tcgaccctct agtcaaggcc ttaagtgagt cgtattacgg actggcgctc gttttacaac 60
gtcgtgactg ggaaccctt ggcgttacct aacttaatcg ccttcgagca catccccctt 120
tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cctttcccaa cagttgcgca 180
gcctgaatgg cgaatggcgc ttcgcttggg aataaagccc gcttcggcgg gctttttttt 240
gttaactacg tcaggtggca cttttcgggg aaatgtgcgc ggaaccccta tttgtttatt 300
tttctaaata cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca 360
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 420
ttttgcggca ttttgccctc ctgtttttgc tcaccagaa acgctgggtg aagtaaaaga 480
tgctgaagat cagttgggtg cagcagtggt ttacatcgaa ctggatctca acagcggtaa 540
gatccttgag agttttcgcc ccgaagaacg ttctccaatg atgagcaact ttaaagtctt 600
gctatgtggc gcggtattat cccgtgttga cgccgggcaa gagcaactcg gtcgcccgat 660
acactattct cagaatgact tggttgagta ctcaccagtc acagaaaagc atcttacgga 720
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata aactgcggc 780
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 840
gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 900
cgacgagcgt gacaccacga tgctgtagc aatggcaaca acgttgcgca aactattaac 960
tggcgaacta cttactctag cttcccggca acaattaata gactggatgg aggcggataa 1020
agttgcagga ccacttctgc gctcggccct tccgctggc tggtttattg ctgataaatc 1080
tggagccggg gagcgtgggt ctcgcggtat cattgcagca ctggggccag atggttaagc 1140
ctcccgtatc gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag 1200
acagatcgct gagtaggtg cctcactgat taagcattgg taactgtcag accaagttta 1260
ctcatatata ctttagattg atttaccocg ttgataatc agaaaagccc caaaaacagg 1320
aagattgtat aagcaaatat ttaaattgta aacgttaata ttttgtaaaa attcgcgtta 1380
aatttttgtt aaatcagctc attttttaac caataggccc aaatcggcaa aatcccttat 1440
aatcaaaaag aatagcccga agtagggttg agtgttggc cagtttggaa caaggtcca 1500
ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc 1560
ccactacgtg aaccatcacc caaatcaagt tttttggggt cgaggtgccg taaagcacta 1620

```

-50-

```

aatcggaacc ctaaagggag cccccgattt agagcttgac ggggaaagcg aacgtggcga 1680
gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcggtca 1740
cgctgcgcggt aaccaccaca cccgcccgcg ttaatgcgcc gctacagggc gcgtaaaagg 1800
atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagtttttcg 1860
ttccactgag cgtcagaccc cgtagaaaaag atcaaaggat cttcttgaga tcctttttttt 1920
ctgcgcgtaaa tctgctgctt gcaaacaataa aaaccaccgc taccagcggg gggttggttg 1980
ccggatcaag agctaccaac tctttttccg aagtaactg gcttcagcag agcgcagata 2040
ccaaataactg ttcttctagt gtagccgtag ttaggccaac acttcaagaa ctctgtagca 2100
ccgcctacat acctcgctct gctaatactg ttaccagtgg ctgctgccag tggcgataag 2160
tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 2220
tgaacggggg gtctgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2280
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagaaaa ggccggacagg 2340
tatccggtaa cggcagggt cggaaacagga gagcgcacga gggagcttcc aggggggaaac 2400
gcctgggtatc tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgattttttg 2460
tgatgctcgt caggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2520
ttcctggcct tttgctggcc ttttgcctac atgtaatgtg agttagctca ctcataggc 2580
acccaggct ttacacttta tgcctccggc tcgtatgttg tgtggaattg tgagcggata 2640
acaatttcac acaggaaaca gctatgacca tgattacgcc aagctacgta atacgactca 2700
ctagtggggc ccgtgcaatt gaagccggct ggcgccaagc ttctctgcag gattgaagcc 2760
tgctttttta tactaacttg agcgaatct ggatccatgg ccaagttgac cagtgcggt 2820
ccgggtgctca ccgcgcgca cgtcgccgga gcggtcgagt tctggaccga ccggtcggg 2880
ttctcccggg acttcgtgga ggacgacttc gccggtgtgg tccgggacga cgtgaccctg 2940
ttcatcagcg cggccagga ccagggtggg ccggaaca cctggcctg ggtgtgggtg 3000
cgcggcctgg acgagctgta ccgagtggtg tcggaggtcg tgtccacgaa cttccgggac 3060
gcctccgggc cggccatgac cgagatcggc gagcagccgt gggggcggga gttcgccctg 3120
cgcgacccgg ccggcaactg cgtgcacttc gtggccgagg agcaggactg acacgtgcta 3180
cgagatttcg attccaccgc cgcttctat gaaaggttgg gcttcggaat cgttttcgg 3240
gacgccggct ggatgatcct ccagcgcggg gatctcatgc tggagtctt cgccccccc 3300
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 3360
aataaagcat tttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 3420
tatcatgtct gtataccg

```

<210> 92
 <211> 10549
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCambia1302 Plasmid

<300>
 <308> Genbank #AF234398
 <309> 2000-04-24

```

<400> 92
catggtagat ctgactagta aaggagaaga acttttcact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaaggtga 120
tgcaacatac ggaaaactta cccttaaat tatttgact actggaaaac tacctgttc 180
gtggccaaca cttgtcacta ctttctctta tgggtgtcaa tgcttttcaa gatacccaga 240
tcatatgaag cggcacgact tcttcaagag cgccatgcct gagggatacg tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacacgt gctgaagtca agtttgagg 360
agacaccctc gtcaacagga tcgagcttaa gggaatcgat ttcaaggagg acggaaacat 420
cctcgccac aagtgggaat acaactacaa ctcccacaa gtatacatca tggccgacaa 480
gcaaaagaac ggcatacaag ccaacttcaa gaccgcacac aacatcgaag acggcgccgt 540
gcaactcgct gatcattatc aacaaaatac tccaattggc gatggccctg tccttttacc 600
agacaaccat tacctgtcca cacaatctgc ccttgcgaaa gatcccaacg aaaagagaga 660
ccacatggtc cttcttgagt ttgtaacagc tgcgtgggatt acacatggca tggatgaact 720
atacaaggct agccaccacc accaccacca cgtgtgaatt ggtgaccagc tcgaatttcc 780
ccgatcgctt aaacatttgg caataaagtt tcttaagatt gaatcctgtt gccggtcttg 840
cgatgattat catataattt ctgttggaat acgttaagca tgtaataatt aacatgtaat 900
gcgatgcgtt atttatgaga tgggttttta tgattagagt cccgcaatta tacatttaat 960
accgcataga tagcgcgcaa tagcggataa attatcgcg attatcgcg gcggtgtcat 1020
ctatgttact agatcgggaa ttaaactatc agtgtttgac aggataatatt ggccgggtaaa 1080
cctaagagaa aagagcgttt attagaataa cggaatattta aaagggcggtg aaaaggttta 1140
tccgttcgtc catttgtagt tgcatgccaa ccacagggtt cccctcgga tcaaagtact 1200
ttgatccaac ccctccgctg ctatagtgc gtcggcttct gacgttcagt gcagccgtct 1260
tctgaaaacg acatgtcgca caagtcctaa gttacgcgac aggtgcccgc cctgcccttt 1320

```


-51-

tcctggcggt	ttcttgtgcg	gtgttttagt	cgcataaagt	agaatacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgccccgcgc	1440
agcaccgacg	accagacttt	gaccaaccaa	cgggcccgaac	tgacgcgggc	cggtctgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccctgg	cgacgtttgtg	acagtgcacca	ggctagaccg	cctggcccg	1620
agcaccgcg	acctaactgga	cattgccgag	cgcatccagg	aggccggcgc	gggcctgcgt	1680
agcctggcag	agcctggggc	cgacaccacc	acgcgcggccg	gcccgcattggt	gttgaccgtg	1740
ttcgccggca	ttgccgagtt	cgagcgttcc	ctaatacatcg	accgcaccgc	gagcggggcg	1800
gaggccgcca	aggcccgagg	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgccgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggctgca	1920
ctgcttggcg	tgcatcgctc	gacctgtac	cgcgacttg	agcgcagcga	ggaagtgcag	1980
cccaccgagg	ccaggcggcg	cggtgccttc	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgggcg	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcaggcgga	gatgatcgcg	gccgggtacg	2160
tggtcgagcc	gcccgcgcac	gtctcaaccg	tgccgctgca	tgaaatcctg	gccggtttgt	2220
ctgatgccaa	gctggcgggc	tgcccgggcca	gcttggccgc	tgaaagaaacc	gagcgccgcg	2280
gtctaaaaag	gtgatgtgta	tttgagttaa	acagcttgcg	tcattgcggtc	gctgcgtata	2340
tgatgcatg	agtaataaaa	caaatacgca	aggggaacgc	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcgccttgca	2460
actcgccggg	ggcgatgttc	tgtagtgcga	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcggccggt	ggcggaagtc	aaccggtaac	cgttgcggc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gccggcgcgca	cttcgtagtg	atcgacggag	cgccccaggc	2640
ggcggacttg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cggtgacgcc	2700
aagcccttca	gacatatggg	ccaccgcgca	cctgggtggag	ctgggttaagc	agcgcattga	2760
ggtcacggat	ggaaggctac	aagcggcctt	tgctcgtgtcg	cgggcatca	aaggcacgcg	2820
catcgccggg	gaggttgccg	aggcgctggc	cgggtacgag	ctgcccattc	ttgagtcgcc	2880
tatcacgcag	cgctgagct	accaggcac	tgccgcgcgc	ggcacaaccg	ttcttgaatc	2940
agaaccgcag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gcccgtgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgcc	3060
ggcgcgtccg	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcca	3120
ggcgttggcc	gggtcaactt	tcagtgtccg	tcaggaggatc	acaccaagct	gaagatgtac	3180
gggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcggctaaag	gaggcggcat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccattgtgtg	gaggaaacggg	3360
cgggttggcca	ggcgtaagcg	gctgggttgt	ctgcccggcc	tgcaatggca	ctggaaacccc	3420
caagcccag	gaatcggcgt	gacggtcgca	aaccatccgg	cccgttaca	atcgccgcgg	3480
cgctgggtga	tgacctgggtg	gagaagtga	aggccgcgca	ggccgcccag	cggcaacgca	3540
tcgaggcgca	agcacgcccc	gggtaacgt	ggcaagcgcc	cgctgatcga	atccgcaaa	3600
aatcccgcca	accgcggcca	gccggtgcgc	cgtcgattag	gaagccgccc	aagggcgacg	3660
agcaaccaga	ttttttcggt	ccgatgctct	atgacgtggg	caccgcgcat	agtcgcagca	3720
tcattggacgt	ggcgttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgaggg	tcacagcggt	cacgtagagg	tttccgcagg	gcccggccggc	atggcgagtg	3840
tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaagggagac	aagcccgccc	gcgtgttccg	tccacacggt	gcggacgtac	3960
tcaagttctg	ccggcgagcc	gatggcgga	gcagaaaaga	cgacctggta	gaaacctgca	4020
ttcggttaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcccgg	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacgggtc	accctgatta	ctttttgatc	gatcccggca	4260
tcggccggtt	tctctaccgc	ctggcacgcc	gcgcccagag	caaggcagaa	gccagatggg	4320
tggtcaagac	gatctacgaa	cgcatgggca	gcgcccggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatgca	cctagcaggg	gaaaaagggtc	4560
gaaaaggctc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagcgc	tacattggga	4620
accggaaccc	gtacattggg	aaccctaaag	cgtacattgg	gaaccggtca	cacatgtaag	4680
tgactgatata	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aacccgcctg	gcctgtgcac	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtcgctgc	gctccctacg	ccccgcgcgt	tcgctcgccg	4860
ctatcgccgg	cgctggccgc	tcaaaaatgg	ctggccctacg	gccaggcaat	ctaccagggc	4920
gcccgaacag	cgccgcgctg	ccactcgacc	gcggcgcccc	acatcaaggc	accctgcctc	4980
gcccgtttcg	gtgatgacgg	tgaaaaacct	tgacacatgc	agctcccggg	gacggtcaca	5040
gcttgcctgt	aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgctc	agcgggtgtt	5100
ggcgggtgtc	ggggcgagc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgc	ggcatcagag	cagattgtac	tgagagtgc	ccatatgcgg	tgtaaatatc	5220
cgcacagatg	cgtaaggaga	aaataccgca	agggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggctcgggc	gagcgggtatc	agctcactca	aaggcggtaa	5340

-52-

tacggttatc	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	gcgaaaccog	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgtgcg	ctctcctggt	ccgaccctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcggtgtagg	tcggtcgctc	caagctgggc	tgtgtgcacg	5700
aacccccgt	tcagcccgac	cgctgcgct	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggatgttagg	cgggtgtaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggcaaacaa	accaccgctg	gtagcgggtg	tttttttgtt	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acggggtctg	6060
acgctcagtg	gaacgaaaaa	tcacgttaag	ggattttggg	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaaata	taattatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatata	ctccctgatc	gaccggacgc	6240
agaagccaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgcc	atctttcaca	aagatgttgc	tgctcccag	gtcgcgctgg	gaaaagacaa	6360
gttctctctc	gggcttttcc	gtctttaaaa	aatcatacag	ctcgcgcgga	tctttaaatg	6420
gagtgctctc	ttcccagttt	tcgcaatcca	catcgccag	atcggtattc	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgattg	6540
agtgaaagag	cctgatgcac	tcgcataaca	cttcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgtccagc	6660
catcatgccc	ttcaaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tcctgcattt	6780
ttaaatatag	gttttcatth	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttoga	tcagtttttt	caattccggt	gatattctca	6900
ttttagccat	ttattatttc	cttccctctt	ttcacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttcttgca	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaaagt	tgttttcaaaa	gttggcgat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcggtaca	atcaacatgc	7140
taccctccgc	agatcatcc	gtgtttcaaaa	cccgccagct	tagttgccgt	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtctgccgc	cttacaacgg	ctctcccgct	gacgccgtcc	7260
cggactgatg	ggctgcctgt	atcgagtggt	gattttgtgc	cgagctgccg	gtcggggagc	7320
tggtggctgg	ctggtggcag	gatataattgt	ggtgtaaaaa	aattgacgct	tagacaactt	7380
aataacacat	tgccgacgtt	tttaattgtac	tgaattaaacg	ccgaattaat	tcgggggatc	7440
tggtatttag	tactggattt	tggttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	aggggtttctt	atatgctcaa	cacatgagcg	aaaccctata	7560
ggaaacctaa	ttcccttata	tgggaaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccggtc	ggcatctact	ctattttctt	gccctcggac	gagtgtcggg	7680
gcgtcgggtt	ccactatcgg	cgagtaacttc	tacacagcca	tcgggtccaga	cggccgcgct	7740
tcgccccggc	atthgtgtac	gccccacagt	cccggtcccg	gatcggacga	ttgctgcga	7800
tcgaccctgc	gccccagctg	catcatcgaa	atthccgtca	accaagctct	gatagagttg	7860
gtcaagacca	atgcccagca	tatacgccc	gagtcgtggc	gatcctgcaa	gctccggatg	7920
cctccgctcg	aagtagcgcg	tctgctgctc	catacaagcc	aaccacggcc	tccagaagaa	7980
gatgttggtg	agctcgatatt	gggaatcccc	gaacatcgcc	tcgctccagt	caatgacccg	8040
tggtatgcgg	ccattgtccg	tcaggacatt	gttgagccg	aaatccgctg	gcacgaggtg	8100
ccggacttcg	gggcagtcct	cggcccaaa	catcagctca	tcgagagcct	gcgcgacgga	8160
cgactgacg	gtgctgcca	tcacagtttg	ccagtgtata	acatggggat	cagcaactcg	8220
gcatatgaaa	tcacgccatg	tagtgtattg	accgatttct	tgcggtccga	atggggccga	8280
cccgtcgtgc	tggttaagat	cggccgcagc	gatcgcaccc	atagcctccg	cgaccgggtg	8340
tagaacagcg	ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccct	gtgcacggcg	8400
ggagatgcaa	taggtcaggg	tctcgctaaa	ctccccaatg	tcaagcactt	cgggaatcgg	8460
gagcgcggcc	gatgcaaagt	gcccataaac	ataacgatct	ttgtagaaac	catcggcgca	8520
gctattttacc	cgcaggacat	atccacgccc	tctacatcg	aagctgaaag	cacgagattc	8580
ttcgccctcc	gagagctgca	tcaggctcga	gacgtgtcg	aacttttcga	tcagaaactc	8640
ctcgacagac	gtcgcgggtg	gttcaggctt	tttcatatct	cattgcccc	cgggatctgc	8700
gaaagctcga	gagagataga	ttttagagaa	gagactggtg	atttcagcgt	gtcctctcca	8760
aatgaaatga	acttccttat	atagaggaag	gtcttgcgaa	ggatagtggtg	attgtgcgtc	8820
atcccttaag	tcagtggaga	tatcacatca	atccacttgc	tttgaagacg	tggttggaac	8880
gtcttctttt	tcacagatgc	tcctcgtggg	tgggggtcca	tctttgggac	cactgtcggc	8940
agagggcatct	tgaacgatag	cctttccttt	atcgcaatga	tggtatttgt	aggtgccacc	9000
ttcctttctc	actgtccttt	tgatgaagtg	acagatagct	gggcaatgga	atccgagga	9060
gtttcccgat	attacccttt	gttgaaaagt	ctcaatagcc	ctttggtctt	ctgagactgt	9120
atctttgata	ttcttgaggt	agacgagagt	gtcgtgctcc	accatgttat	cacatcaatc	9180
cacttgcttt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	tcgtgggtgg	9240
gggtccatct	ttgggaccac	tgctcggcaga	ggcatcttga	acgatagcct	ttcctttatc	9300
gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	tgaagtgaca	9360

-53-

gatagctggg	caatggaatc	cgaggaggtt	tcccgatatt	accctttggt	gaaaagtctc	9420
aatagccctt	tggctctctg	agactgtatc	tttgatattc	ttggagtaga	cgagagtgtc	9480
gtgctccacc	atgttggcaa	gctgctctag	ccaatacgca	aaccgcctct	ccccgcgcgt	9540
tggccgattc	attaatgcag	ctggcacgac	aggtttcccg	actggaaagc	gggcagttag	9600
cgcaacgcaa	ttaatgtgag	ttagctcact	cattagggac	cccaggcttt	acactttatg	9660
cttccggctc	gtatgttggt	tgggaattgt	agcggataac	aatttcacac	aggaaacagc	9720
tatgacctag	attacgaatt	cgagctcgg	accgggggat	cctctagagt	cgacctgcag	9780
gcatgcaagc	ttggcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accctggcgt	9840
tacccaactt	aatcgccttg	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	9900
ggccgcgacc	gatcgccctt	cccaacagtt	gcgcagcctg	aatggcgaat	gctagagcag	9960
cttgagcttg	gatcagattg	tcgtttcccg	ccttcagttt	agcttcatgg	agtcaaagat	10020
tcaaatagag	gacctaacag	aactcgccgt	aaagactggc	gaacagttca	tacagagtct	10080
cttaccgactc	aatgacaaga	agaaaatctt	cgtcaacatg	gtggagcacg	acacacttgt	10140
ctactccaaa	aatatcaaa	atacagctc	agaagaccaa	agggcaattg	agacttttca	10200
acaaagggtta	atatccggaa	acctcctcgg	attccattgc	ccagctatct	gtcactttat	10260
tgtgaagata	gtggaaaagg	aaggtggctc	ctacaaatgc	catcattgcg	ataaaggaaa	10320
ggccatcggt	gaagatgcct	ctgccgacag	tggtcccaaa	gatggacccc	cacccacgag	10380
ggcatcgtg	gaaaaagaag	acgttccaac	cacgtcttca	aagcaagtgg	attgatgtga	10440
tatctccact	gacgtaagg	atgacgcaca	atccactat	ccttcgcaag	acccttcctc	10500
tatataagga	agttcatttc	atttggagag	aacacggggg	actcttgac		10549

<210> 93

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> CaMV35SPolyA Primer

<400> 93

ctgaattaac gccgaattaa ttccgggggat ctg

33

<210> 94

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> CaMV35Spr Primer

<400> 94

ctagagcagc ttgccaacat ggtggagca

29

<210> 95

<211> 12592

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg2 Plasmid

<400> 95

gtacgaagaa	ggccaagaac	ggccgcctgg	tgacgggtatc	cgagggtgaa	gccttgatta	60
gccgctacaa	gatcgtaaa	agcgaaccg	ggcggccgga	gtacatcgag	atcgagctag	120
ctgattggat	taccgcgag	atcacagaag	gcaagaaccc	ggacgtgctg	acggttcacc	180
ccgattactt	tttgatcgat	cccggcatcg	gccgttttct	ctaccgcctg	gcacgcccg	240
ccgcaggcaa	ggcagaagcc	agatggttgt	tcaagacgat	ctacgaacgc	agtggcagcg	300
ccggagagtt	caagaagttc	tgtttcaccg	tgcgcaagct	gatcgggtca	aatgacctgc	360
cggagtacga	tttgaaggag	gaggcggggc	aggctggccc	gatcctagtc	atgcgctacc	420
gcaacctgat	cgagggcgaa	gcatccgccg	gttcctaattg	tacggagcag	atgctagggc	480
aaattgccct	agcaggggaa	aaaggtcgaa	aaaggtctctt	tcctgtggat	agcacgtaca	540
ttgggaaccc	aaagccgtac	attgggaacc	ggaaccogta	cattgggaac	ccaaagccgt	600
acattgggaa	ccggtcacac	atgtaagtga	ctgatataaa	agagaaaaaa	ggcgattttt	660
ccgcctaaaa	ctctttaaaa	cttattaaaa	ctcttaaaac	ccgcctggcc	tgtgcataac	720
tgtctggcca	gcgcacagcc	gaagagctgc	aaaaagcgcc	tacccttcgg	tcgctgcgct	780
ccctacgccc	cgccgcttcg	cgccgctcta	tcgcggccgc	tggccgctca	aaaatggctg	840
gcctacggcc	aggcaatcta	ccagggcgcg	gacaagccgc	gccgtcgcca	ctcgaccgcc	900

-54-

ggcgccca	tcaaggcacc	ctgcctcgcg	cgtttcggtg	atgacgggtga	aaacctctga	960
cacatgcagc	tcccggagac	ggtcacagct	tgtctgtaag	cggatgcccgg	gagcagacaa	1020
gcccgtcagg	gcgcgtcagc	gggtgttggc	gggtgtcggg	gcgcagccat	gacccaggtca	1080
cgtagcgata	gcccagtgta	tactggotta	actatgcggc	atcagagcag	attgtactga	1140
gagtgaccca	tatgcggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	1200
ggcgctcttc	cgcttcctcg	ctcactgact	cgctgcgctc	ggtcgttcgg	ctgcggcgag	1260
cggtatcagc	tcactcaaag	gcggtaatac	ggttatccac	agaatcaggg	gataacgcag	1320
gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcggtgc	1380
tggcgttttt	ccataggctc	cgccccctcg	acgagcatca	caaaaaatcga	cgctcaagtc	1440
agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	1500
tcgtgcgctc	tcctgttccg	accctgccgc	ttaccggata	cctgtccgcc	tttctccctt	1560
cgggaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	gtgtagggtcg	1620
ttcgctccaa	gctgggctgt	gtgcacgaac	ccccggtca	gcccagccgc	tgcgcttat	1680
ccggtaaacta	tcgtcttgag	tcacaacccg	taagacacga	cttatcgcca	ctggcagcag	1740
ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	ttcttgaagt	1800
ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	1860
cagttacctt	cggaaaaaaga	gttggtagct	cttgcgctcg	caaaacaaacc	accgctggta	1920
gcggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaag	1980
atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	cgaaaactca	cgtaaagggga	2040
ttttggatcat	gcatcttagg	tactaaaaca	attcatccag	taaaatataa	tattttattt	2100
tctcccaatc	aggcttgatc	cccagtaagt	caaaaaatag	ctcgacatac	tggtcttccc	2160
cgatatcctc	cctgatcgac	cggacgcaga	aggcaatgtc	ataccacttg	tccgcctctgc	2220
cgcttctccc	aagatcaata	aagccactta	ctttgcccac	tttcacaaag	atgttgctgt	2280
ctcccaggtc	gccgtgggaa	aagacaagtt	cctcttcggg	cttttccgctc	tttaaaaaat	2340
catacagctc	gcgcggatct	ttaaatggag	tgtcttcttc	ccagttttcg	caatccacat	2400
cggccagatc	gttattcagt	aagtaatcca	attcggctaa	gcggctgtct	aagctattcg	2460
tatagggaca	atccgatatg	tcgatggagt	gaaagagcct	gatgcactcc	gcatacagct	2520
cgataatctt	ttcagggtct	tggtcatctt	catactcttc	cgagcaaagg	acgccatcgg	2580
cctcactcat	gagcagattg	ctccagccat	catgccgttc	aaagtgcagg	acctttggga	2640
caggcagctt	tccttccagc	catagcatca	tgtccttttc	ccgttccaca	tcatagggtg	2700
tcctcttata	ccgctgttcc	gtcattttta	aatatagggt	ttcattttct	cccaccagct	2760
tatatacctt	agcaggagac	attccttccg	tatcttttac	gcagcgggat	ttttcgatca	2820
gttttttcaa	ttccgggtgat	attctcattt	tagccattta	ttatttccct	cctcttttct	2880
acagttattt	aagatacccc	aagaagctaa	ttataacaag	acgaactcca	attcactggt	2940
ccttgcattc	taaaacctta	aataccagaa	aacagctttt	tcaaagttgt	tttcaaagtt	3000
ggcgtataac	atagtatcga	cggagccgat	tttgaaaccg	cgggtgatcac	aggcagcaac	3060
gctctgtcat	cgttacaatc	aacatgctac	cctccgcgag	atcatccgtg	tttcaaaccc	3120
ggcagcttag	ttgcggttct	tcogaatagc	atcggttaaca	tgagcaaaagt	ctgcgcgctt	3180
acaacggctc	tcccgtgac	gccgtcccgg	actgatgggc	tgctgtatc	gagtggtgat	3240
tttgtgccga	gctgccggtc	ggggagctgt	tggctggctg	gtggcaggat	atattgtggt	3300
gtaaacaaat	tgagccttag	acaacttaat	aacacattgc	ggacgttttt	aatgtactga	3360
attaacgcgc	aattaattcg	ggggatctgg	attttagtag	tggatttttg	tttttaggaat	3420
tagaaatttt	attgatagaa	gtattttaca	aatacaata	cataactaagg	gtttcttata	3480
tgctcaacac	atgagcgaaa	ccctatagga	accctaattc	ccttatctgg	gaactactca	3540
cacattatta	tgagaaaact	cgagtgacg	ctcgtgacg	ggcaggaccg	gacggggcgg	3600
taccggcagg	ctgaagtcca	gctgccagaa	acccacgtca	tgccagttcc	cgtgcttgaa	3660
gccggccgcc	cgcagcatgc	cgcggggggc	atatccgagc	gcctcgtgca	tgccgcacgt	3720
cgggtcgltg	ggcagcccga	tgacagcgac	cacgctcttg	aagccctgtg	cctccaggga	3780
cttcagcagg	tgggtgtaga	gcgtggagcc	cagtcctctg	cgctgggtggc	ggggggagac	3840
gtacacgggtc	gactcggccg	tccagtcgta	ggcgttgctg	gccttccagg	ggcccgcgta	3900
ggcgatgccg	gcgacctcgc	cgtccacctc	ggcgacgagc	cagggatagc	gctccgcgag	3960
acggagcgagg	tcgtccgtcc	actcctgcgg	ttcctgcggc	tcggtacgga	agttgacggt	4020
gcttgtctcg	atgtagtggg	tgacgatggg	gcagaccgcc	ggcatgtccg	cctcgggtggc	4080
acggcgggatg	tcggccgggc	gtcgttctgg	gctcatggta	gactcgagag	agatagattt	4140
gtagagagag	actggtgatt	tcagcgtgtc	ccttccaaat	gaaatgaact	tccttatata	4200
gaggaagggtc	ttgcgaagga	tagtgggatt	gtgcgtcatc	ccttacgtca	gtggagatat	4260
cacatcaatc	cacttgcttt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	4320
tcgtgggtgg	gggtccatct	ttgggaccac	tgctcggcaga	ggcatcttga	acgatagcct	4380
ttccttttatc	gcaatgatgg	cattttgtagg	tgccaccttc	cttttctact	gtccttttga	4440
tgaagtgaca	gatagctggg	caatggaatc	cgaggagggtt	tcccgatatt	accctttggt	4500
gaaaagtctc	aatagccctt	tggtcttctg	agactgtatc	tttgatatctc	ttggagtaga	4560
cgagtgtgct	gtgctccacc	atgttatcac	atcaatccac	ttgctttgaa	gacgtgggtg	4620
gaacgtcttc	tttttccacg	atgctcctcg	tggtgggggg	tccatctttg	ggaccactgt	4680
cggcagaggc	atcttgaacg	atagcctttc	ctttatcgca	atgatggcat	ttgtaggtgc	4740
caccttccct	tctactgtc	cttttgatga	agtgacagat	agctgggcaa	tggaaatcca	4800
ggagggtttcc	cgatattacc	ctttgttgaa	aagcttcaat	agccctttgg	tcttctgaga	4860
ctgtatcttt	gatattcttg	gagtagacga	gagtgctcgtg	ctccaccatg	ttggcaagct	4920

gctctagcca	atacgcaaac	cgccctctccc	cgcgcggttg	ccgattcatt	aatgcagctg	4980
gcacgcacag	tttcccgact	ggaaagcggg	cagtgcgcgc	aacgcatta	atgtgagtta	5040
gctcactcat	taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tggtgtgtgg	5100
aattgtgagc	ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgaattcga	5160
gccttgacta	gagggctcgac	ggtatacaga	catgataaga	tacattgatg	agtttgga	5220
aaccacaact	agaatgcagt	gaaaaaaatg	ctttatttgt	gaaatttgtg	atgctattgc	5280
tttatttgtg	accattataa	gctgcaataa	acaagttggg	gtgggcgaag	aactccagca	5340
tgagatcccc	gcgctggagg	atcatccagc	cggcgtcccc	gaaaacgatt	ccgaagccca	5400
acctttcata	gaaggcggcg	gtggaatcga	aatctcgtag	cacgtgtcag	tcctgtctct	5460
cgccacgaa	gtgcacgcag	ttgccggccg	ggtcgcgcag	ggcgaactcc	cgccccacg	5520
gctgctcgcc	gatctcggtc	atggccggcc	cgaggcgctc	ccggaagttc	gtggacacga	5580
cctccgacca	ctcggcgctac	agctcgtcca	ggccgcgcac	ccacacccag	gccagggtgt	5640
tgctccggcac	cacctggctc	tggaccgcgc	tgatgaacag	ggtcacgtcg	tcccggacca	5700
caccggcgaa	ctcgtctccc	acgaagtccc	gggagaaccc	gagcgggtcg	gtccagaact	5760
cgaccgctcc	ggcgacgtcg	cgcgcggtga	gcaccggaac	ggcactgggtc	aacttggcca	5820
tggatccaga	tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	caggaattcg	5880
atcgacactc	tcgtctactc	caagaatata	aaagatacac	tctcagaaga	ccaaagggtt	5940
attgagactt	ttcaacaaag	ggtaatatcg	ggaaacctcc	tcggattcca	ttgccagct	6000
atctgtcact	tcatacaaaag	gacagtagaa	aaggaaggtg	gcacctacaa	atgccatcat	6060
tgcgataaag	gaaaggctat	cgttcaagat	gcctctgcgc	acagtgggtc	caaagatgga	6120
ccccaccca	cgaggagcat	cgtggaaaaa	gaagacgttc	caaccacgtc	ttcaagcaa	6180
gtggattgat	gtgataacat	ggtggagcac	gacactctcg	tctactccaa	gaatatcaaa	6240
gatacagtct	cagaagacca	aagggtat	gagacttttc	aacaaagggt	aatatcgga	6300
aacctcctcg	gattccattg	cccagctatc	tgctcacttc	tcaaaaggac	agtagaaaag	6360
gaagggtggc	ctacaaaatg	ccatcattgc	gataaaggaa	aggctatcgt	tcaagatgcc	6420
tctgccgaca	gtggtcccaa	agatggaccc	ccaccacaga	ggagcatcgt	ggaaaaagaa	6480
gacgttccaa	ccacgtcttc	aaagcaagtg	gattgatgtg	atatctccac	tgacgtaagg	6540
gatgacgcac	aatcccacta	tccttcgcaa	gaccttcctc	tatataagga	agttcatttc	6600
atttgagag	gacacgctga	aatcaccagt	ctctctctac	aaatctatct	ctctcgagct	6660
ttcgagatc	cgggggggc	atgagatag	aaaaagcctg	aactcaccgc	gacgtctgtc	6720
gagaagtttc	tgatcgaaaa	gttcgacagc	tgatccgacc	tgatgcagct	ctcggagggc	6780
gaagaatctc	gtgctttcag	cttcgatgta	ggagggcggtg	gatatgtcct	gcgggtaaat	6840
agctgcgcgc	atgggtttcta	caaagatcgt	tatgtttatc	ggcactttgc	atcgcccgcg	6900
ctcccagatt	cggaagtgt	tgacattggg	gagtttagcg	agagcctgac	ctattgcac	6960
tcccgcctg	cacagggtgt	cacgttgcaa	gacctgcctg	aaaccgaact	gcccgtgtt	7020
ctacaaccgg	tcgcgagggc	tatggatgcg	atcgctgcgc	ccgatcttag	ccagacgagc	7080
gggttccggc	cattcgggac	gcaaggaaatc	gggtcaataca	ctacatggcg	tgatttcata	7140
tgcgcgattg	ctgatcccca	tggtgtatcac	tggtcaaaactg	tgatggacga	caccgtcagt	7200
gcgtccgtcg	cgcaggctct	cgatgagctg	atgctttggg	ccgaggactg	ccccgaagtc	7260
cggcacctcg	tgacgcggga	tttcggctcc	aacaatgtcc	tgacggacaa	tggccgcata	7320
acagcgggtc	ttgactggag	cgaggcgatg	ttcggggatt	cccaatacga	ggtcgccaac	7380
atcttcttct	ggaggccgtg	gttggttgt	atggagcagc	agacgcgcta	cttcgagcgg	7440
aggcatccgg	agcttgagg	atcgccacga	ctccggcggt	atagctccg	cattgggtctt	7500
gaccaactct	atcagagctt	ggttgacggc	aatttcgatg	atgcagcttg	ggcgagggtt	7560
cgatgcgacg	caatcgctcg	atccggagcg	ggcgtagctg	ggcgtagaca	aatcgcccg	7620
agaagcgcg	ccgtctggac	cgatggctgt	gtagaagtac	tcgccgatag	tggaaaccga	7680
cgccccagca	ctcgctccgag	ggcaaaagaa	tagagtagat	gccgaccgga	tctgtcgatc	7740
gacaagctcg	agttttctcca	taataatgtg	tgagttagttc	ccagataagg	gaattagggt	7800
tctataggg	tttcgctcat	gtgttgagca	tataagaaac	ccttagtatg	tatttgatt	7860
tgtaaaatac	ttctatcaat	aaaatttcta	attcctaaaa	ccaaaatcca	gtactaaaat	7920
ccagatcccc	cgaattaatt	cggcgttaat	tcagatcaag	cttggcactg	gccgtcgttt	7980
tacaacgtcg	tgactgggaa	aaccctggcg	ttaccctaact	taatcgctt	gcagcacatc	8040
cccccttctc	cagctggcgt	aatagcgaag	aggcccgac	cgatcgccct	tcccaacagt	8100
tgccgagcct	gaatggcgaa	tgctagagca	gcttgagctt	ggatcagatt	gtcgtttccc	8160
gccttcagtt	tggggatcct	ctagactgaa	ggcgggaaac	gacaatctga	tcattgagcgg	8220
agaattaagg	gagtcacgtt	atgaccccg	ccgatgacgc	gggacaagcc	gttttacgtt	8280
tggaaactgac	agaaccgcaa	cggtgaagga	gccactcagc	cgcggttttc	tggagtttaa	8340
tgagctaagc	acatacgtca	gaaaccatta	ttgcgcgttc	aaaagtgcgc	taaggctcact	8400
atcagtagc	aaatatctct	tgtcaaaaaat	gctccactga	cgttccataa	attccccctg	8460
gtatccaatt	agagtctcat	attcactctc	aatccaaata	atctgcaccg	gatctcgaga	8520
atcgaattcc	cgcggccgc	atggttagatc	tgactagtaa	aggagaagaa	cttttctactg	8580
gagttgtccc	aattcttgtt	gaattagatg	gtgatgttaa	tgggcacaaa	ttttctgtca	8640
gtggagaggg	tgaaggtgat	gcaacatacg	gaaaacttac	ccttaaattt	atttgcacta	8700
ctggaaaact	acctgttccg	tggccaacac	ttgtcactac	tttctcttat	ggtgttcaat	8760
gcttttcaag	ataccagat	catatgaagc	ggcacgactt	cttcaagagc	gccatgcctg	8820
agggatacgt	gcaggagagg	accatcttct	tcaaggacga	cggggaactac	aagacagctg	8880
ctgaagtcaa	gtttgagggg	gacacectcg	tcaacaggat	cgagcttaag	ggaatcgatt	8940

-56-

tcaaggagga	cggaacatc	ctcggccaca	agttggaata	caactacaac	tcccacaacg	9000
tatacatcat	ggcgcgacaag	caaaagaacg	gcatacaaac	caacttcaag	acccgccaca	9060
acatcgaaga	cgccggcgtg	caactcgctg	atcattatca	acaaaatact	ccaattggcg	9120
atggccctgt	cctttttacca	gacaaccatt	acctgtccac	acaatctgcc	ccttcgaaag	9180
atcccaacga	aaagagagac	cacatggtcc	ttcttgagtt	tgtaacagct	gctgggatta	9240
cacatggcat	ggatgaacta	tacaaagcta	gccaccacca	ccaccaccac	gtgtgaattg	9300
gtgaccagct	cgaattttccc	cgatcgttca	aacatttgge	aataaagttt	cctaagattg	9360
aatcctgttg	ccggtccttgc	gatgattatc	atataatttc	tggtgaatta	cgtaagcat	9420
gtaataatta	acatgtaatg	catgacgtta	tttatgagat	gggtttttat	gattagagtc	9480
ccgcaattat	acatttaata	cgcgatagaa	aacaaaatat	agcgcgcaaa	ctaggataaa	9540
ttatcgcgcg	cggtgtcatc	tatgttacta	gatcggaat	taaactatca	gtgtttgaca	9600
ggatatattg	gcgggtaaac	ctaagagaaa	agagcggtta	ttagaataac	ggatatatta	9660
aagggcgtga	aaaggtttat	ccgttcgtcc	atthgtatgt	gcatgccaac	cacagggttc	9720
ccctcgggat	caaagtactt	tgatccaacc	cctccgctgc	tatagtgcag	tcggcttctg	9780
acgttcagtg	cagccgtctt	ctgaaaacga	catgtcgcac	aagtcctaag	ttacgcgaca	9840
ggctgcgcgc	ctgccttttt	cctggcggtt	tcttgcgcgc	tgthtttagtc	gcataaagta	9900
gaataacttg	gactagaacc	ggagacatta	cgccatgaac	aagagcgccg	ccgctggcct	9960
gtggggctat	gcccgcgtca	gcaccgacga	ccaggacttg	accaaccaac	gggcccgaact	10020
gcacgcggcc	ggctgcacca	agctgttttc	cgagaagatc	accggcacca	ggcgcgaccg	10080
cccggagctg	gccaggatgc	ttgaccacct	acgcccctggc	gacgttgtga	cagtgaccag	10140
gctagaccgc	ctggcccgca	gcaccccgca	cctactggac	attgccgagc	gcacccagga	10200
ggccggcgcg	ggcctgcgta	gcctggcaga	gccgtgggccc	gacaccacca	cgccggcccg	10260
ccgcatgggtg	ttgaccgtgt	tcgcccggcat	tgccgagttc	gagcgttccc	taatcatcga	10320
ccgcaccgcg	agcgggcgcg	aggccgcgaa	ggcccgaggc	gtgaagtttg	gcccccgccc	10380
taccctcacc	gcggcacaga	tcgcgacgcg	ccgcgagctg	atcgaccagg	aaggccgcac	10440
cgtgaaagag	gcggctgcac	tgcttgccgt	gcatacgctc	accctgtacc	gcgcacttga	10500
gcgcagcgag	gaagtgcgcg	ccaccgaggc	caggcggcgc	ggtgccttcc	gtgaggacgc	10560
attgaccgag	gccgacgccc	tgccggccgc	cgagaatgaa	cgccaagagg	aacaagcatg	10620
aaaccgcacc	aggacggcca	ggacgaaccg	tttttcatta	ccgaagagat	cgaggcggag	10680
atgatcgcg	ccgggtacgt	gttcgagccg	cccgcgcacg	tctcaaccgt	gcccgtgcac	10740
gaaatcctgg	ccggtttgtc	tgatgccaa	ggccggccct	ggccggccag	cctggcctct	10800
gaagaaaccg	agcgcgcgcg	tctaaaaagg	tgatgtgtat	ttgagtaaaa	cagcttgctg	10860
catgcggctc	ctgcgtatat	gatgcgatga	gtaaataaac	aaatacgcaa	ggggaacgca	10920
tgaaggttat	cgctgtactt	aaccagaaag	gcgggtcagg	caagacgacc	atcgcaacc	10980
atctagcccg	ccgcttgcaa	ctgcgcgggg	ccgatgttct	gttagtcgat	tccgactccc	11040
agggcagtg	ccgcgattgg	gcggccgtgc	gggaagatca	accgctaacc	gthgtcgcca	11100
tcgaccgccc	gacgatggac	cgcgacgtga	aggccatcgg	ccggcgcgac	ttcgtagtga	11160
tcgctggagc	gcccagggcg	gcggacttgg	ctgtgtccgc	gatcaaggca	gccgacttgc	11220
tgctgattcc	ggtgcagcca	agcccttacg	acatatgggc	caccgcccgc	ctggtggagc	11280
tggttaagca	gcgcattgag	gtcacggatg	gaaggctaca	agcggccttt	gtcgtgtcgc	11340
ggggcatcaa	aggcacgcgc	atcggcggtg	aggttgccga	ggcgtggccc	gggtacgagc	11400
tgccatttct	tgagtcccgt	atcacgcagc	gcgtgagcta	cccaggcact	gccgcgcgcg	11460
gcacaaccgt	tcttgaatca	gaacccgagg	gcgacgctgc	ccgcgagggtc	caggcgctgg	11520
ccgctgaaat	taaatcaaaa	ctcatttgag	ttaatgaggt	aaagagaaaa	tgagcaaaa	11580
cacaaacacg	ctaagtgcgc	gccgtccgag	cgcacgcgagc	agcaaggctg	caacgttggc	11640
cagcctggca	gacacgcccag	ccatgaagcg	ggtcaacttt	cagttgcccgc	cggaggatca	11700
caccaagctg	aagatgtacg	cggtacgcca	aggcaagacc	attaccgagc	tgctatctga	11760
atacatcgcg	cagctaccag	agtaaatgag	caaatgaata	aatgagtaga	tgaatttttag	11820
cggttaaagg	aggcggcatg	gaaaatcaag	aacaaccagg	caccgacgcc	gtggaatgcc	11880
ccatgtgtgg	aggaaacgggc	ggttggccag	gcgtaagcgc	ctgggttgct	tgccggccct	11940
gcaatggcac	tggaaaccccc	aagcccagg	aatcggcggtg	acggctcgcaa	accatccggc	12000
ccggtacaaa	tcggcgcggc	gctgggtgat	gacctgggtg	agaagttgaa	ggccgcgcag	12060
gccgcccagc	ggcaacgcat	cgaggcagaa	gcacgccccg	gtgaatcgtg	gcaagcggcc	12120
gctgatcgaa	tccgcaaaga	atcccggcaa	ccgcccggcag	ccggtgcgcgc	gtcgattagg	12180
aagccgcccc	aggcgacga	gcaaccagat	tttttcgttc	cgatgctcta	tgacgtgggc	12240
acccgcgata	gtcgcagcat	catggacgtg	gcccgttttcc	gtctgtcgaa	gcgtgaccga	12300
cgagctggcg	aggtgatccg	ctacgagctt	ccagacgggc	acgtagaggt	ttccgcaggg	12360
ccggccggca	tggccagtg	gtgggattac	gacctggtac	tgatggcggt	ttcccatcta	12420
accgaatcca	tgaaccgata	ccgggaagg	aagggagaca	agcccggccg	cgtgttccgt	12480
ccacacgttg	cggacgtact	caagttctgc	cggcgagccg	atggcggaaa	gcagaaagac	12540
gacctggtag	aaacctgcat	tcggttaaac	accacgcacg	ttgccatgca	gc	12592

<210> 96

<211> 3357

<212> DNA

<213> Artificial Sequence

-57-

<220>

<223> pGEMEasyNOS Plasmid

<400> 96

tatcactagt	gaattcgcgg	ccgcctgcag	gtcgaccata	tgggagagct	cccaacgcgt	60
tggatgcata	gcttgagtat	tctatagtgt	cacctaata	gcttggcgta	atcatggtca	120
tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacattc	cacacaacat	acgagccgga	180
agcataaagt	gtaaagcctg	gggtgcctaa	tgagtgcct	aactcacatt	aattgctgtg	240
cgctcactgc	ccgctttcca	gtcgggaaac	ctgtcgtgcc	agctgcatta	atgaatcggc	300
caacgcgcgg	ggagaggcgg	tttgcgtatt	gggcgctctt	ccgcttcctc	gctcactgac	360
tcgctgcgct	cggctcgttcg	gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtaata	420
cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	480
aaggccagga	accgtaaaaa	ggccgcgttg	ctggcgtttt	tccataggct	ccgccccct	540
gacgagcatc	acaaaaatcg	acgctcaagt	cagagggtgc	gaaacccgac	aggactataa	600
agataccagg	cgtttccccc	tggaagctcc	ctcgtgcgct	ctcctgttcc	gacctgcccg	660
cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	720
cgctgaaggt	atctcagttc	ggtgtaggte	gttcgctcca	agctgggctg	tgtgcacgaa	780
ccccccgttc	agcccgaccg	ctgcgcctta	tccggttaact	atcgtcttga	gtccaacccg	840
gtaagacacg	acttatcgcc	actggcagca	gccactggta	acaggattag	cagagcgagg	900
tatgtaggcg	gtgctacaga	gttcttgaag	tggtggccta	actacggcta	cactagaaga	960
acagtatttg	gtattctgcg	tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	1020
tcttgatccg	gcaaaacaaac	caccgctggt	agcgggtggt	tttttgtttg	caagcagcag	1080
attacgcgca	gaaaaaaagg	atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	1140
gctcagtgga	acgaaaaactc	acgttaaggg	attttggtca	tgagattatc	aaaaaggatc	1200
ttcacctaga	tcctttttaa	ttaaaaatga	agttttaaat	caatctaaag	tatatatgag	1260
taaacttggt	ctgacagtta	ccaatgctta	atcagtgagg	cacctatctc	agcgatctgt	1320
ctatttcggt	catccatagt	tgctgactc	cccgtcgtgt	agataactac	gatacgggag	1380
ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	acccaagctc	accgggtcca	1440
gatttatcag	caataaacca	gccagccgga	agggccgagc	gcagaagtgg	tcctgcaact	1500
ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	ctagagtaag	tagttcgcca	1560
gttaataagt	tgccgaacgt	tggtgccatt	gctacaggca	tcgtgggtgc	acgctcgctg	1620
tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	ggcgagttac	atgatcccc	1680
atggttggtca	aaaaagcggg	tagctccttc	gggtcctccga	tcgttgctcag	aagtaagttg	1740
gccgcagtg	tatcactcat	ggttatggca	gcactgcata	attctcttac	tgtcatgcca	1800
tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	agtcattctg	agaatagttg	1860
atgcggcgac	cgagttgctc	ttgcccggcg	tcaatacggg	ataataccgc	gccacatagc	1920
agaactttaa	aagtgtcat	cattggaaaa	cgttcttcgg	ggcgaaaaact	ctcaaggatc	1980
ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	atcttcagca	2040
tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	gaaggcaaaa	tgccgcacaaa	2100
aagggaataa	gggcgacacg	gaaatgttga	atactcatac	tcttcccttt	tcaatattat	2160
tgaagcattt	atcagggtta	ttgtctcatg	agcggataca	tatttgaatg	tatttagaaa	2220
aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	tgccggtgtga	2280
aataccgcac	agatgcgtaa	ggagaaaata	ccgcatacagg	aaattgtaag	cgttaatat	2340
ttgttaaaat	tcgcgttaaa	tttttgtaaa	atcagctcat	tttttaacca	ataggccgaa	2400
atcgccaaaa	tcccttataa	atcaaaagaa	tagaccgaga	taggggttag	tgttgttcca	2460
gtttggaaca	agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	2520
gtctatcagg	gcgatggccc	actacgtgaa	ccatcacctc	aatcaagttt	tttgggggtcg	2580
agggtgcgta	aagcactaaa	tcggaacctc	aaaggagacc	cccgatttag	agcttgacgg	2640
ggaaaagcgg	cgaacgtggc	gagaaaaggaa	gggaagaaaag	cgaaaggagc	gggcgctagg	2700
gcgctggcaa	gtgtagcggg	cacgctgcgc	gtaaccacca	caccgcgcgc	gcttaatgcg	2760
ccgctacagg	gcgcgtccat	tcgccattca	ggctgcgcaa	ctgttgggaa	ggcgcatcgg	2820
tgccggcctc	ttcgctatta	cgccagctgg	cgaaaaggggg	atgtgctgca	aggcgattaa	2880
gttgggtaac	gccagggttt	tcccagtcac	gacgttgtaa	aacgacggcc	agtgaattgt	2940
aatacgactc	actatagggc	gaattggggc	cgacgtcgca	tgctcccggc	cgccatggcg	3000
gccgcgggaa	ttcgattctc	gagatccggt	gcagattatt	tggattgaga	gtgaatatga	3060
gactctaatt	ggataccgag	gggaatttat	ggaacgtcag	tggagcattt	ttgacaagaa	3120
atatttgcta	gctgatagtg	accttagggc	acttttgaac	gcgcaataat	ggtttctgac	3180
gtatgtgctt	agctcattaa	actccagaaa	cccgcggctg	agtggctcct	tcaacgttgc	3240
ggttctgtca	gttccaaacg	ttaaaccggc	tgtcccgcgt	catcgccggg	ggtcataaac	3300
tgactccctt	aattctccgc	tcatgatcag	attgtcgttt	ccgccttca	gtctaga	3357

<210> 97

<211> 10122

<212> DNA

<213> Artificial Sequence

<220>

-58-

<223> p1302NOS Plasmid

<400> 97

```

catggttagat ctgactagta aaggagaaga acttttctact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaaggtga 120
tgcaacatac ggaaaactta cccttaaat ttttgcact actggaaaac tacctgttcc 180
gtggccaaca cttgtcacta ctttctotta tggtgttcaa tgcttttcaa gataccaga 240
tcatatgaag cggcacgact tcttcaagag cgccatgcct gagggatacg tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacacgt gctgaagtca agtttgaggg 360
agacaccctc gtcaacagga tcgagcttaa gggaatcgat ttcaaggagg acggaacat 420
cctcggccac aagttggaat acaactacaa ctcccacaac gtatacatca tggccgacaa 480
gcaaaagaac ggcatacaag ccaacttcaa gaccgcacac aacatcgaag acggcggcgt 540
gcaactcgct gatcattatc acaaaaatag tccaattggc gatggccctg tccttttacc 600
agacaacccat tacctgtcca cacaatctgc cctttcgaaa gatcccaacg aaaagagaga 660
ccacatgggtc cttcttgagt ttgtaacagc tgctgggatt acacatggca tggatgaact 720
atacaaaagct agccaccacc accaccacca cgtgtgaatt ggtgaccagc tcgaatttcc 780
ccgatcgctt aaacatttgg ctgttgaatt acgttaagca tgtaataatt, aacatgtaat 900
cgatgattat atttatgaga tgggttttta tgattagagt cccgcaatta tacattttaa 960
acgcgataga aaacaaaata tagcgcgcaa actaggataa attatcgcg gcggtgtcat 1020
ctatgttact agatcgggaa ttaaaactac agtgtttgac aggatataatt ggcggttcaa 1080
cctaagagaa aagagcggtt attagaataa cggatattta aaagggcggtg aaaaggttta 1140
tccgttcgtc catttgtatg tgcatgccaa ccacagggtt cccctcgga tcaaagtact 1200
ttgatccaac cctcgcgtg ctatagtga caagtcctaa gttacgcgac gagcttcagt gcagcgtct 1260
tctgaaaacg acatgtcgca gtgttttagt cgcataaagt ggcgctggcc aggtgcccgc cctgcccttt 1320
tcctggcggt ttcttgtcgc caagagcgcc ggcgctggcc tgcacgcggc cactagaaac 1380
cggagacatt accgcatgaa gaccaaccaa aggcgcgacc ggctagaccg cctggcccgc 1440
agcaccgacg accaggactt cagcgtttgt cgcacgcggc aggcgcggcg ggcaggatg 1500
aagctgtttt cgcgagagat cagccctggg cgttgcggag acagtgacca ggcgcgagct cctggcccgc 1560
cttgaccacc tacgcccctg cgttgcggag cgttgcggag aggcgcggcg ggcgcgagct cctggcccgc 1620
agcaccgcg agcctggcag agcctggcgg cgtgaagttt ggcccccgcc ctacctcac cccggcacag 1680
agcctggcag ttgcccaggt agcccgtggc cgtgaagttt ggcccccgcc ctacctcac cccggcacag 1740
ttcgcggcga ttgcccaggt agcccgtggc cgtgaagttt ggcccccgcc ctacctcac cccggcacag 1800
gagggcgcca agggccgagg cgtgaagttt ggcccccgcc ctacctcac cccggcacag 1860
atcgcgcagc cccgcgagct gatcgaccag gaagggcgca cgtgaaagta ggcgcgagct cccggcacag 1920
ctgcttggcg tgcacgcctc cccgcgagct cgtgaagttt ggcccccgcc ctacctcac cccggcacag 1980
cccaccgagg ccaggcgagg cgtgaagttt ggcccccgcc ctacctcac cccggcacag 2040
ctggcgggcg ccaggcgagg gatcgaccag gaaaccgcac caggaaggcg ggcgggtacg 2100
aggacgaacc gcccgcgcac gtttttcatt accgaagaga tgcaggcgga gatgatcgcg gcccgggtacg 2160
tggtcgcgac gctggcgggc tttgagtaaa acagcttgcg tgaagaaacc gccggtttgt 2220
tgatgcgatg agtaataaaa caaatcgca aggggaacgc atgaaggtta tgcgctgtact 2280
taaccagaaa ggcgggtcag gcaagacgac catcgcaacc catctagccc gcgccctgca 2340
actcgcggg cccgatgttc cgggaagatc gctgtgtccg ccccgctgca gctgctgata 2400
ggcgggcggt cgggaagatc gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 2460
ccgcgacgtg aaggccatcg gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 2520
aagcccttac gacataggg ccaccgcca gctgtgtccg ccccgctgca gctgctgata 2580
ggtcacggat ggaaggctac gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 2640
catcgcggt gaggttgccg aggcgtggc ccccgctgca gctgctgata tgcgctgtact 2700
tatcacgcag cgcgtgagct acccaggcac tgcgcggc ccccgctgca gctgctgata 2760
agaaccgag ggcgacgctg gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 2820
actcatttga gttaatgagg gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 2880
ggcgtccga ggcacgcag gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 2940
gccatgaagc ggtcaactt aaggcaagac gctgtgtccg ccccgctgca gctgctgata 3000
gcggtacgcc aaggcaagac gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3060
gagtaaatga gcaaatgaat gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3120
ggaaaatcaa gaacaaccag gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3180
cgggttgcca ggcgtaagcg gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3240
caagcccag gactcggcgt gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3300
cgctgggtga tgacctgggt gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3360
tcgagcgaga agcacgcccc gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3420
aatcccggca accgcggcca gctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3480
agcaaccaga ttttttctgt cctgtgtccg ccccgctgca gctgctgata tgcgctgtact 3540
tcattggcgt tccagacggg cactagaggg tttccgcagg gcccggccggc agtgcgacga 3600
gctacgagct tccagacggg cactagaggg tttccgcagg gcccggccggc agtgcgacga 3660

```


tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaagggagac	aagcccgcc	gcgtgttcgg	tccacacgtt	gcggacgtac	3960
tcaagtcttg	ccggcgagcc	gatggcggaa	agcagaaaaga	cgacctggta	gaaacctgca	4020
ttcgggttaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcccgc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	ccgggacgtg	ctgacgggtc	accccgatta	ctttttgatc	gatcccgcga	4260
tccggccgtt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggg	4320
tggtcaagac	gatctacgaa	cgcaagtggc	gcgcgcggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	ccgatcccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccgggttccta	atgtacggag	cagatgctag	ggcaaattgc	cctagcaggg	gaaaaagggtc	4560
gaaaagggtc	gtttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaacct	gtacattggg	aaccocaaag	cgtacattgg	gaaccgggtc	cacatgtaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aacctgcctg	gcctgtgcat	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggctgcgtc	gctccctacg	ccccgcgcgt	tcgcgtcggc	4860
ctatcgccgg	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcggacaagc	cgccgcgctc	ccactcgacc	gccggcgccc	acatcaaggc	acctgcctc	4980
gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacggtcaca	5040
gcttgctctg	aagcggtatg	cgggagcgag	caagcccgct	agggcgcgct	agcgggtggt	5100
ggcgggtgtc	ggggcgacgc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttaactatgc	ggcatcagag	cagattgtac	tgagagtgca	ccatatgcgg	tgtgaaatac	5220
cgacacatgc	ctaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctcgc	ctcggtcggt	cggtgcggc	gagcggtatc	agctcactca	aaggcggtaa	5340
tacgggttat	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggcccgct	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	gcgaaaccgc	acaggactat	5520
aaagatacca	ggcggtttcc	cctggaagct	ccctcgctcg	ctctcctggt	ccgacctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtag	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aacccccctg	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggtaaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggatatgtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagga	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagtgggta	5940
gctcttgatc	cggaacacaa	accaccgctg	gtagcgggtg	tttttttggt	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccctt	gatcttttct	acggggtctg	6060
acgctcagtg	gaacgaaaac	tcacgttaag	ggattttggt	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaata	taatatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatatc	ctccctgatc	gaccggacgc	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgcc	atctttcaca	aagatgttgc	tgctcccgag	gtcgccgtgg	gaaaagacaa	6360
gttccctctc	gggcttttcc	gtctttaaaa	aatcatacag	ctcgccggga	tctttaaatg	6420
gagtgctctc	ttcccgagtt	tcgcaatcca	catcgccag	atcggtattc	agtaagtaat	6480
ccaattccgg	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tccgcataca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgctccagc	6660
catcatgccg	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcct	ttcccgcttc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttcattt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tattttttcg	tcagtttttt	caattccggg	gatatttctc	6900
tttttagccat	ttattatttc	cttcctcttt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttccttgca	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcgat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcggttaca	atcaacatgc	7140
tacctccgc	gagatcatcc	cccggcagct	cccggcagct	tagttgccc	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtctgccgc	cttacaacgg	ctctcccgc	gacgcggtcc	7260
cggactgatg	ggctgcctgt	atcgagtggg	gattttgtgc	cgagctgccg	gtcggggagc	7320
tggtggctgg	ctgggtggcag	gatataattg	gggtgtaaca	aattgacgct	tagacaactt	7380
aataacacat	tgccgacggt	tttaattgtac	tgaattaacg	ccgaattaat	tcgggggatc	7440
tggattttag	tactggattt	tgggttttag	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgctcaa	cacatgagcg	aaaccttata	7560
ggaaccctaa	ttcccttata	tggaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccggtc	ggcatctact	ctatttcttt	gccctcggac	gagtgtcggg	7680
gcgtcgggtt	ccattatcgg	cgagtaactc	tacacagcca	tcgggtccaga	cggccgcgct	7740
tctcgggcgg	acttgtgtac	gcccgacagt	cccggtccg	gacggacga	ttgcgtcgca	7800
tcgacctgc	gccccagctg	catcatcgaa	attgccgtca	accaagctct	gatagagttg	7860

-60-

```

gtcaagacca atgcccagca tatacgcggc gagtcgtggc gatcctgcaa gctccggatg 7920
cctccgctcg aagtagcgcg tctgctgctc catacaagcc aaccacggcc tccagaagaa 7980
gatgttggcg acctcgctatt gggaatcccc gaacatcgcc tcgctccagt caatgaccgc 8040
tggtatgocg ccattgtccg tcaggacatt gttggagccg aaatccgcgt gcacgaggtg 8100
ccggacttcg gggcagtcct cggcccaaaag catcagctca tcgagagcct gcgcgacgga 8160
cgcactgacg gtgtcgtcca tcacagtttg ccagtatac acatggggat cagcaatcgc 8220
gcatatgaaa tcacgcatg tagtgtattg accgattcct tgcgggtccga atgggcccga 8280
cccgcctcgtc tggctaagat cggccgcagc gatcgcatcc atagcctccg cgaccggttg 8340
tagaacagcg ggcagttcgg tttcaggcag gtcttgcaac gtgacaccct gtgcacggcg 8400
ggagatgcaa taggtcaggg tctcgctaaa ctccccaatg tcaagcactt ccggaatcgg 8460
gagcgcggcc gatgcaaagt gccgataaac ataacgatct ttgtagaaac catcggcgca 8520
gctattttacc cgcaggacat atccacgccc tccacatcg aagctgaaag cacgagattc 8580
ttcgcctccc gagagctgca tcaggtcggg gacgtgtcg aacttttcga tcagaaactt 8640
ctcgacagac gtgcgggtga gttcaggctt tttcatatct cattgcccc ccggtatctg 8700
gaaagctcga gagagataga tttgtagaga gagactggtg atttcagcgt gtccctccca 8760
aatgaaatga acttcccttat atagaggaag gtcttcgcaa ggatagtggg attgtgcgtc 8820
atcccttacg tcagtgagga tatcacatca atccacttgc tttgaagacg ttggttgaac 8880
gtcttctttt tccacgatgc tctcgtggg tgggggtcca tctttgggac cactgtcggc 8940
agaggcatct tgaacgatag cctttccttt atcgcaatga tggcatttgt aggtgccacc 9000
ttccttttct actgtccttt tgatgaagtg acagatagct gggcaatgga atccgaggag 9060
gtttcccgat attacccttt gttgaaaagt ctcaatagcc ctttgggtct ctgagactgt 9120
atctttgata ttcttgaggt agacgagagt gtcgtgctcc accatgttat cacatcaatc 9180
cacttgcctt gaagacgtgg ttggaacgtc ttctttttcc acgatgctcc tcgtgggtgg 9240
gggtccatct ttgggaccac tgcggcaga ggcactctga acgatagcct ttcctttatc 9300
gcaatgatgg catttgtagg tgccaccttc cttttctact gtccctttga tgaagtgaac 9360
gatagctggg caatggaatc cgaggagggt tcccgatatt accctttgtt gaaaaagtct 9420
aatagccctt tggctctctg agactgtatc tttgatattc ttggagtaga cgagagtgtc 9480
gtgctccacc atgttggcaa gctgctctag ccaatacgca aaccgcctct cccgcgcgt 9540
tgcccgattc attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag 9600
cgcaacgcaa ttaattgtgag ttagctcact cattaggcac cccaggcttt acactttatg 9660
cttcgggtc gtatgttggt tggaaattgt agcgataaac aatttcacac aggaaacagc 9720
tatgaccatg attacgaatt cgagctcggg acccggggat cctctagact gaaggcggga 9780
aacgacaatc tgatcatgag cggagaatta agggagtac gttatgacct ccgccgatga 9840
cgccgggaaa gccgttttac gtttggaaat gacagaaccg caacgttgaa ggagccactc 9900
agccgcgggt ttctggagtt taatgagcta agcacatacg tcagaaacca ttattgcgcg 9960
ttcaaaagtc gcctaagggt actatcagct agcaaatatt tcttgtcaaa aatgctccac 10020
tgacgttcca taaattcccc tcggtatcca attagagtct catattcact ctcaatccaa 10080
ataatctgca ccggtatctc agaatcgaat tcccgcggcc gc 10122

```

<210> 98
 <211> 621
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> N. tabacum rDNA intergenic spacer (IGS) sequence

<300>
 <308> Genbank #Y08422
 <309> 1997-10-31

```

<400> 98
gtgctagcca atgtttaaca agatgtcaag cacaatgaat gttggtggtt ggtggtcgtg 60
gctggcggtg gtggaaaatt gcggtggttc gagcggtagt gatcggcgat ggttgggtgt 120
tgcagcggtg tttgatatcg gaatcactta tgggtggtgt cacaatggag gtgcgtcatg 180
gttattgggtg gttgggtcatc tatatatttt tataataata ttaagtattt tacctatttt 240
ttacatattt tttattaaat ttatgcattg tttgtatttt taaatagttt ttatcgtact 300
tgttttataa aatattttat tattttatgt gttatattat tacttgatgt attggaaatt 360
ttctccattg ttttttctat atttataata attttcttat ttttttttgt ttattatgt 420
attttttcgt tttataataa atatttatta aaaaaaatat tatttttgta aaatatatca 480
tttacaatgt ttaaaagtca tttgtgaata tattagctaa gttgtacttc tttttgtgca 540
tttgggtgtg tacatgtcta ttatgattct ctggccaaaa catgtctact cctgtcactt 600
gggttttttt ttttaagaca t

```

<210> 99
 <211> 25
 <212> DNA

-61-

<213> Artificial Sequence

<220>

<223> NTIGS-F1 Primer

<400> 99

gtgctagcca atgtttaaca agatg

25

<210> 100

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-R1 Primer

<400> 100

atgtcttaaa aaaaaaaacc caagtgc

28

<210> 101

<211> 233

<212> DNA

<213> Mus Musculus

<300>

<308> Genbank #V00846

<309> 1989-07-06

<400> 101

gacctggaat atggcgagaa aactgaaaat cacggaaaat gagaaataca cacttttagga 60
 cgtgaaatat ggcgaggaaa actgaaaaag gtggaaaatt tagaaatgtc cactgtagga 120
 cgtggaatat ggcaagaaaa ctgaaaatca tggaaaatga gaaacatcca cttgacgact 180
 tgaaaaatga cgaaatcact aaaaaacgtg aaaaatgaga aatgcacact gaa 233

<210> 102

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-F1 Primer

<400> 102

aataccgcgg aagcttgacc tggaatatcg c

31

<210> 103

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-R1 Primer

<400> 103

ataaccgcgg agtccttcag tgtgcat

27

<210> 104

<211> 277

<212> DNA

<213> Artificial Sequence

<220>

<223> Nopaline Synthase Promoter Sequence

<300>

<308> Genbank #U09365

<309> 1997-10-17

-62-

```

<400> 104
gagctcgaat ttccccgatc gttcaaacaat ttggcaataa agttttcttaa gattgaatcc 60
tggtgccggt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cgttatattat gagatgggtt tttatgatta gagtcccgca 180
attatacatt taatacgcga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240
gcgcgcggtg tcatttatgt tactagatcg ggaattc 277

```

```

<210> 105
<211> 1812
<212> DNA
<213> Escherichia coli

```

```

<220>
<221> CDS
<222> (1)...(1812)
<223> Beta-Glucuronidase

```

```

<300>
<308> Genbank #S69414
<309> 1994-09-23

```

```

<400> 105
atg tta cgt cct gta gaa acc cca acc cgt gaa atc aaa aaa ctc gac 48
Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
1 5 10 15

ggc ctg tgg gca ttc agt ctg gat cgc gaa aac tgt gga att gat cag 96
Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
20 25 30

cgt tgg tgg gaa agc gcg tta caa gaa agc cgg gca att gct gtg cca 144
Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
35 40 45

ggc agt ttt aac gat cag ttc gcc gat gca gat att cgt aat tat gcg 192
Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
50 55 60

ggc aac gtc tgg tat cag cgc gaa gtc ttt ata ccg aaa ggt tgg gca 240
Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
65 70 75 80

ggc cag cgt atc gtg ctg cgt ttc gat gcg gtc act cat tac ggc aaa 288
Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
85 90 95

gtg tgg gtc aat aat cag gaa gtg atg gag cat cag ggc ggc tat acg 336
Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
100 105 110

cca ttt gaa gcc gat gtc acg ccg tat gtt att gcc ggg aaa agt gta 384
Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
115 120 125

cgt atc acc gtt tgt gtg aac aac gaa ctg aac tgg cag act atc ccg 432
Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
130 135 140

ccg gga atg gtg att acc gac gaa aac ggc aag aaa aag cag tct tac 480
Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
145 150 155 160

ttc cat gat ttc ttt aac tat gcc gga atc cat cgc agc gta atg ctc 528
Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
165 170 175

tac acc acg ccg aac acc tgg gtg gac gat atc acc gtg gtg acg cat 576

```

-63-

Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185					190			
gtc	gcg	caa	gac	tgt	aac	cac	gcg	tct	gtt	gac	tgg	cag	gtg	gtg	gcc	624
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala	
		195					200					205				
aat	ggt	gat	gtc	agc	gtt	gaa	ctg	cgt	gat	gcg	gat	caa	cag	gtg	gtt	672
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val	
	210					215					220					
gca	act	gga	caa	ggc	act	agc	ggg	act	ttg	caa	gtg	gtg	aat	ccg	cac	720
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His	
225					230					235					240	
ctc	tgg	caa	ccg	ggg	gaa	ggg	tat	ctc	tat	gaa	ctg	tgc	gtc	aca	gcc	768
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala	
				245					250					255		
aaa	agc	cag	aca	gag	tgt	gat	atc	tac	ccg	ctt	cgc	gtc	ggc	atc	cgg	816
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg	
			260					265					270			
tca	gtg	gca	gtg	aag	ggc	gaa	cag	ttc	ctg	att	aac	cac	aaa	ccg	ttc	864
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe	
		275					280					285				
tac	ttt	act	ggc	ttt	ggg	cgt	cat	gaa	gat	gcg	gac	ttg	cgt	ggc	aaa	912
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys	
	290					295					300					
gga	ttc	gat	aac	gtg	ctg	atg	gtg	cac	gac	cac	gca	tta	atg	gac	tgg	960
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp	
305					310					315					320	
att	ggg	gcc	aac	tcc	tac	cgt	acc	tcg	cat	tac	cct	tac	gct	gaa	gag	1008
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu	
				325					330					335		
atg	ctc	gac	tgg	gca	gat	gaa	cat	ggc	atc	gtg	gtg	att	gat	gaa	act	1056
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr	
			340					345					350			
gct	gct	gtc	ggc	ttt	aac	ctc	tct	tta	ggc	att	ggg	ttc	gaa	gcg	ggc	1104
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly	
		355					360					365				
aac	aag	ccg	aaa	gaa	ctg	tac	agc	gaa	gag	gca	gtc	aac	ggg	gaa	act	1152
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr	
		370				375					380					
cag	caa	gcg	cac	tta	cag	gcg	att	aaa	gag	ctg	ata	gcg	cgt	gac	aaa	1200
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys	
385					390					395					400	
aac	cac	cca	agc	gtg	gtg	atg	tgg	agt	att	gcc	aac	gaa	ccg	gat	acc	1248
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr	
				405					410					415		
cgt	ccg	caa	ggg	gca	cgg	gaa	tat	ttc	gcg	cca	ctg	gcg	gaa	gca	acg	1296
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr	
			420					425					430			
cgt	aaa	ctc	gac	ccg	acg	cgt	ccg	atc	acc	tgc	gtc	aat	gta	atg	ttc	1344
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe	
		435					440					445				

-64-

tgc gac gct cac acc gat acc atc agc gat ctc ttt gat gtg ctg tgc 1392
 Cys Asp Ala His Thr Asp Thr Ile Ser Asp Leu Phe Asp Val Leu Cys
 450 455 460

ctg aac cgt tat tac gga tgg tat gtc caa agc ggc gat ttg gaa acg 1440
 Leu Asn Arg Tyr Tyr Gly Trp Tyr Val Gln Ser Gly Asp Leu Glu Thr
 465 470 475 480

gca gag aag gta ctg gaa aaa gaa ctt ctg gcc tgg cag gag aaa ctg 1488
 Ala Glu Lys Val Leu Glu Lys Glu Leu Leu Ala Trp Gln Glu Lys Leu
 485 490 495

cat cag ccg att atc atc acc gaa tac ggc gtg gat acg tta gcc ggg 1536
 His Gln Pro Ile Ile Ile Thr Glu Tyr Gly Val Asp Thr Leu Ala Gly
 500 505 510

ctg cac tca atg tac acc gac atg tgg agt gaa gag tat cag tgt gca 1584
 Leu His Ser Met Tyr Thr Asp Met Trp Ser Glu Glu Tyr Gln Cys Ala
 515 520 525

tgg ctg gat atg tat cac cgc gtc ttt gat cgc gtc agc gcc gtc gtc 1632
 Trp Leu Asp Met Tyr His Arg Val Phe Asp Arg Val Ser Ala Val Val
 530 535 540

ggt gaa cag gta tgg aat ttc gcc gat ttt gcg acc tcg caa ggc ata 1680
 Gly Glu Gln Val Trp Asn Phe Ala Asp Phe Ala Thr Ser Gln Gly Ile
 545 550 555 560

ttg cgc gtt ggc ggt aac aag aaa ggg atc ttc act cgc gac cgc aaa 1728
 Leu Arg Val Gly Gly Asn Lys Lys Gly Ile Phe Thr Arg Asp Arg Lys
 565 570 575

ccg aag tcg gcg gct ttt ctg ctg caa aaa cgc tgg act ggc atg aac 1776
 Pro Lys Ser Ala Ala Phe Leu Leu Gln Lys Arg Trp Thr Gly Met Asn
 580 585 590

ttc ggt gaa aaa ccg cag cag gga ggc aaa caa tga 1812
 Phe Gly Glu Lys Pro Gln Gln Gly Gly Lys Gln *
 595 600

<210> 106
 <211> 603
 <212> PRT
 <213> Escherichia coli

<300>
 <308> Genbank #S69414
 <309> 1994-09-23

<400> 106
 Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
 1 5 10 15
 Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
 20 25 30
 Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
 35 40 45
 Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
 50 55 60
 Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
 65 70 75 80
 Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
 85 90 95
 Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
 100 105 110
 Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
 115 120 125

-65-

```

Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
130 135 140
Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
145 150 155 160
Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
165 170 175
Tyr Thr Thr Pro Asn Thr Trp Val Asp Asp Ile Thr Val Val Thr His
180 185 190
Val Ala Gln Asp Cys Asn His Ala Ser Val Asp Trp Gln Val Val Ala
195 200 205
Asn Gly Asp Val Ser Val Glu Leu Arg Asp Ala Asp Gln Gln Val Val
210 215 220
Ala Thr Gly Gln Gly Thr Ser Gly Thr Leu Gln Val Val Asn Pro His
225 230 235 240
Leu Trp Gln Pro Gly Glu Gly Tyr Leu Tyr Glu Leu Cys Val Thr Ala
245 250 255
Lys Ser Gln Thr Glu Cys Asp Ile Tyr Pro Leu Arg Val Gly Ile Arg
260 265 270
Ser Val Ala Val Lys Gly Glu Gln Phe Leu Ile Asn His Lys Pro Phe
275 280 285
Tyr Phe Thr Gly Phe Gly Arg His Glu Asp Ala Asp Leu Arg Gly Lys
290 295 300
Gly Phe Asp Asn Val Leu Met Val His Asp His Ala Leu Met Asp Trp
305 310 315 320
Ile Gly Ala Asn Ser Tyr Arg Thr Ser His Tyr Pro Tyr Ala Glu Glu
325 330 335
Met Leu Asp Trp Ala Asp Glu His Gly Ile Val Val Ile Asp Glu Thr
340 345 350
Ala Ala Val Gly Phe Asn Leu Ser Leu Gly Ile Gly Phe Glu Ala Gly
355 360 365
Asn Lys Pro Lys Glu Leu Tyr Ser Glu Glu Ala Val Asn Gly Glu Thr
370 375 380
Gln Gln Ala His Leu Gln Ala Ile Lys Glu Leu Ile Ala Arg Asp Lys
385 390 395 400
Asn His Pro Ser Val Val Met Trp Ser Ile Ala Asn Glu Pro Asp Thr
405 410 415
Arg Pro Gln Gly Ala Arg Glu Tyr Phe Ala Pro Leu Ala Glu Ala Thr
420 425 430
Arg Lys Leu Asp Pro Thr Arg Pro Ile Thr Cys Val Asn Val Met Phe
435 440 445
Cys Asp Ala His Thr Asp Thr Ile Ser Asp Leu Phe Asp Val Leu Cys
450 455 460
Leu Asn Arg Tyr Tyr Gly Trp Tyr Val Gln Ser Gly Asp Leu Glu Thr
465 470 475 480
Ala Glu Lys Val Leu Glu Lys Glu Leu Leu Ala Trp Gln Glu Lys Leu
485 490 495
His Gln Pro Ile Ile Ile Thr Glu Tyr Gly Val Asp Thr Leu Ala Gly
500 505 510
Leu His Ser Met Tyr Thr Asp Met Trp Ser Glu Glu Tyr Gln Cys Ala
515 520 525
Trp Leu Asp Met Tyr His Arg Val Phe Asp Arg Val Ser Ala Val Val
530 535 540
Gly Glu Gln Val Trp Asn Phe Ala Asp Phe Ala Thr Ser Gln Gly Ile
545 550 555 560
Leu Arg Val Gly Gly Asn Lys Lys Gly Ile Phe Thr Arg Asp Arg Lys
565 570 575
Pro Lys Ser Ala Ala Phe Leu Leu Gln Lys Arg Trp Thr Gly Met Asn
580 585 590
Phe Gly Glu Lys Pro Gln Gln Gly Gly Lys Gln
595 600

```

<210> 107

<211> 277

<212> DNA

<213> Artificial Sequence

-66-

<220>

<223> Nopaline Synthase Terminator Sequence

<300>

<308> U09365

<309> 1995-10-17

<400> 107

gagctcgaat	ttccccgcatc	gttcaaacat	ttggcaataa	agttttcttaa	gattgaatcc	60
tgttgccggt	cttgcgatga	ttatcatata	atttctgttg	aattacgtta	agcatgtaat	120
aattaacatg	taatgcgatga	cgttatttat	gagatgggtt	tttatgatta	gagtccccgca	180
attatacatt	taatacgcga	tagaaaacaa	aatatagcgc	gcaaactagg	ataaattatc	240
gcgcgcggtg	tcattctatgt	tactagatcg	ggaattc			277

<210> 108

<211> 3451

<212> DNA

<213> Artificial Sequence

<220>

<223> HindIII Fragment containing the beta-glucuronidase coding sequence, the rDNA intergenic spacer, and the Mast1 sequence

<400> 108

aagcttgacc	tggaatatcg	cgagtaaact	gaaaatcacg	gaaaatgaga	aatacacact	60
ttaggacgtg	aaatatggcg	aggaaaactg	aaaaaggtgg	aaaatttaga	aatgtccact	120
gtaggacgtg	gaatatggca	agaaaaactga	aaatcatgga	aaatgagaaa	catccacttg	180
acgacttgaa	aaatgacgaa	atcactaaaa	aacgtgaaaa	atgagaaatg	cacactgaag	240
gactccgcgg	gaattcgatt	gtgctagcca	atgtttaaca	agatgtcaag	cacaatgaat	300
gttggtgggt	ggtggtcggtg	gctggcggtg	gtggaaaatt	gcggtgggtc	gagcggtagt	360
gatcgcgcat	ggttggtggt	tgcagcggtg	tttgatatcg	gaatcactta	tggtgggtgt	420
cacaatggag	gtgcgctcatg	gttattgggtg	gttggtcatc	tatatatttt	tataataata	480
ttaagtattt	tacctatttt	ttacataattt	tttattaaat	ttatgcattg	tttgtatttt	540
taaaatagttt	ttatcgtaact	tgttttataa	aatattttat	tattttatgt	gttatattat	600
tacttgatgt	attggaaaatt	ttctccattg	ttttttctat	atttataata	attttcttat	660
ttttttttgt	tttattatgt	attttttcgt	tttataataa	atattttatta	aaaaaaatat	720
tattttttgta	aaatatatca	tttacaatgt	ttaaaagtca	tttgtgaata	tattagctaa	780
gttggtacttc	ttttttgtgca	tttggtgttg	tacatgtcta	ttatgattct	ctggccaaaa	840
catgtctact	cctgtcactt	gggttttttt	ttttaagaca	taatcactag	tgattatata	900
tagactgaag	gcgggaaacg	acaatctgat	catgagcgga	gaattaaggg	agtcacgtta	960
tgacccccgc	cgatgacgcg	ggacaagcgc	ttttacgttt	ggaactgaca	gaaccgcaac	1020
gttgaaggag	ccactcagcc	gcgggtttct	ggagtttaat	gagctaagca	catacgtcag	1080
aaaccattat	tgcgcgttca	aaagtgcgct	aagggtcacta	tcagctagca	aatattttctt	1140
gtcaaaaatg	ctccactgac	gttccataaa	ttcccctcgg	tatccaatta	gagtcctata	1200
ttcactctca	atccaaataa	tctgcaccgg	atctcgagat	cgaattcccg	cggccgcgaa	1260
ttcactagtg	gatccccggg	tacggtcagt	cccttatgtt	acgtcctgta	gaaaccccaa	1320
cccgtgaaat	caaaaaactc	gacggcctgt	gggcattcag	tctggatcgc	gaaaactgtg	1380
gaattgagca	gcgttggtgg	gaaagcgcgt	tacaagaaag	ccgggcaatt	gctgtgccag	1440
gcagttttta	cgatcagttc	gccgatgcag	atattcgtaa	ttatgtgggc	aacgtctggt	1500
atcagcgcga	agtctttata	ccgaaaaggt	gggcaggcca	gcgtatcgtg	ctgcgtttcg	1560
atgcgggtcac	tcattacggc	aaagtgtggg	tcaataatca	ggaagtgatg	gagcatcagg	1620
gcgggtatatac	gccatttgaa	gccgatgtca	cgccgtatgt	tattgcccgg	aaaagtgtac	1680
gtatcacagt	ttgtgtgaac	aacgaaactga	actggcagac	tatcccgcgc	ggaatggtga	1740
ttaccgacga	aaacggcaag	aaaaagcagt	cttacttcca	tgattttctt	aactacgccg	1800
ggatccatcg	cagcgtaatg	ctctacacca	cgccgaacac	ctgggtggac	gatatacccg	1860
tggtgacgca	tgctgcgcaa	gactgtaacc	acgcgtctgt	tgactggcag	gtggtggcca	1920
atggtgatgt	cagcgttgaa	ctgcgtgatg	cggatcaaca	ggtggttgca	actggacaag	1980
gcaccagcgg	gactttgcaa	gtggtgaatc	cgcacctctg	gcaaccgggt	gaaggttatc	2040
tctatgaact	gtacgtcaca	gccaaaagcc	agacagagtg	tgatatctac	ccgctgcgcg	2100
tccgcatccg	gtcagtgcca	gtgaagggcg	aacagttcct	gatcaaccac	aaaccgttct	2160
actttactgg	ctttggccgt	catgaagatg	oggttttgcg	cggcaaagga	ttcgataacc	2220
tgctgatggg	gcacgatcac	gcattaatgg	actggattgg	ggccaactcc	taccgtacct	2280
cgcattaccc	ttacgctgaa	gagatgctcg	actgggcaga	tgaacatggc	atcgtggtga	2340
ttgatgaaac	tgacagctgtc	ggctttaacc	tctctttagg	cattgggttc	gaagcgggca	2400
acaagccgaa	agaactgtac	agcgaagagg	cagtcacagg	ggaaactcag	caggcgact	2460
tacaggcgat	taaagagctg	atagcgcgtg	acaaaaacca	cccaagcgtg	gtgatgtgga	2520

-67-

```

gtattgccaa cgaaccggat acccgtccgc aaggtgcacg ggaatatttc gcgccactgg 2580
cggaagcaac gcgtaaaactc gatccgacgc gtccgatcac ctgcgtcaat gtaatgttot 2640
gcgacgctca caccgatacc atcagcgatc tctttgatgt gctgtgcctg aaccgttatt 2700
acggttggtg tgtccaaagc ggcgatttgg aaacggcaga gaaggtactg gaaaaagaac 2760
ttctggcctg gcaggagaaa ctgcatcagc cgattatcat caccgaatac gcggtggata 2820
cgttagccgg gctgcactca atgtacaccg acatgtggag tgaagagtat cagtgtgcat 2880
ggctggatat gtatcaccgc gtctttgata gcgtcagcgc cgtcgtcggg gaacaggtat 2940
ggaatttcgc cgatttttgc acctcgcaag gcatatttgc cgttggcggg aacaagaagg 3000
ggatcttcac ccgcgaccgc aaaccgaagt cggcggcttt tctgctgcaa aaacgctgga 3060
ctggcatgaa cttcgggtgaa aaaccgcagc agggaggcaa acaatgaatc aacaactctc 3120
atcgtcggct acagcctcgg gaatttgcgt ccgagctcga atttcccccga 3180
tcgttcaaac atttggcaat aaagtttctt aagattgaat cctgttgccg gtcttgccgat 3240
gattatcata taattttctg tgaattacgt taagcatgta ataattaaca tgtaatgcata 3300
gacgttatatt atgagatggg tttttatgat tagagtcccg caattataca tttaatacgc 3360
gatagaaaac aaaatatagc gcgcaaaacta ggataaatta tcgcgcgcgg tgatcatctat 3420
gttactagat cgggaattcg atatcaagct t 3451

```

<210> 109

<211> 14627

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg11a Plasmid

<400> 109

```

catgccaaac acaggggttcc cctcgggatc aaagtacttt gatccaaccc ctccgctgct 60
atagtgcagt cggcttctga cgttcagtgc agccgtcttc tgaaaacgac atgtcgcaca 120
agtcctaagt tacgcgacag gctgccgccc tgcccttttc ctggcgtttt ctgtgcgct 180
gttttagtcg cataaagtag aatacttgcg actagaaccg gagacattac gccatgaaca 240
agagcgcgcg cgctggcctg ctgggctatg cccgcgtcag caccgacgac caggacttga 300
ccaaccaacg ggccgaactg cagcggcgcc gctgcaccaa gctgttttcc gagaagatca 360
ccggcaccag gcgcgaccgc ccggagctgg ccaggatgct tgaccaccta cgccctggcg 420
acgttgtgac agtgaccagg ctagaccgcc tggcccgcag caccgcgac ctactggaca 480
ttgccgagcg catccaggag gccggcgccg cctgctgtag cctggcagag cgtgggccc 540
acaccaccac gccggccggc cgcattggtg tgaccgtgtt cgccggcatt gccgagttcg 600
agcgttccct aatcatcgac cgcacccgga gccggcgccg ggcgcgcaag gcccgaggcg 660
tgaagtttgg ccccgccctt accctcaccc gtgaaagagg cggctgcact gcgcacgccc cgcagctga 720
tcgaccaggg aggcgcacc cgacattgag cgacgcgagg aagtgcgccc caccgaggcc catcgctcga 780
ccctgtaccg tgaggacgca ttgaccgagg ccgacgcctt ggccggccgc gagaatgaac 840
gtcccttccg tgagagcatg aaccgcacca cggacggccag gacgaaccgt ttttcattac 900
gccaagagga acaagcatga gaggcggaga tgatcgcgcc cgggtacgtg ttogagccgc ccgcgcacgt 1020
ctcaaccgtg cggctgcatg aaatcctggc cggtttgtct gatgccaaagc tggcgccctg 1080
ggcggccagc ttggccgctg agcttgctgc atgctgctgc ctataaaagg gatgtgtatt 1140
tgagtaaaac ggggaacgcat atgctgctgc gctgtactta accagaaagg taaataaaca 1200
aatacgcaag tgcgaaccca tctagccccg gccctgcaac tcgccggggc cggttcaggc 1260
aagacgacca ccgatcccca ggccagtgcc cgcgattggg cggccgtgcg ggaagatcaa 1320
ttagtcgatt ccgatcccca cgaccgcccg acgattgacc cggccgtgcg ggccatcgcc 1380
ccgctaaccg ttgtcgccat cgacggagcg ccccgaggcg cggacttggc tgtgtcccg 1440
cggcgcgact tcgtagtgat cgtgattccg ggttaagcag ggcacgcgca tcggcgggtg 1500
atcaagtcag ccgacttcgt gctgattccg ggttaagcag ggcacgcgca tcggcgggtg 1560
accgccgacc tgggtggagc tgcgtgctgc ggcgatcaaa gagtcccgtg tccgcgagcg 1620
gcggcccttg ggtacgagct ccgcgcgcgg cactgaatcag aacccgaggg tcatattgag 1680
ccaggcactg ccgcgcgcgg cgctgaaatt aaatcaaaac ccttccgagc gcacgcagca 1740
cgcgaggtcc gagcaaaaag aacgttggcc agcctggcag acacgccagc catgaagcgg 1800
aagagaaaat aacgttggcc agcctggcag agatgtacgc ggtacgcca gtaaatgagc 1860
gcaaggtgct ggaggatcac gctatctgaa ggctaaaagg ggcggcatgg ggaacgggag 1920
agttgcccgc ttaccgagct gaatttttag catgtgtgga caatggcact ggaaccccca 1980
atgagtagat tgggaatgcc gccggccctg ccgtacaaa cggcgcgccg gaggcagaag 2040
accgacggcg tgggttgtct ccatccggcc ccgcccagcg cgcgcaagaa tcccggcaac 2100
tgagttgaag caagcggccc ctgatcgaat cgcgcaagaa 2160

```

-68-

cggtgocgocg	tgcattagga	agcgcgccaa	gggcgacgag	caaccagatt	ttttcggtcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tgcgagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggtt	tccgcagggc	cgcccgccat	ggccagtgtg	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cgggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgagggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggcccg	agtacatcga	3000
gacgagctta	cgtgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgcct	3120
ggcacgcgcg	gcccgcaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	ccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtacg	atltgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tccgagggcg	agcatccgcc	ggttcctaat	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgccttc	gcgtcgccct	atcgcgcccg	ctggccgctc	3720
aaaaatggct	ggcctaagcc	caggcaatct	accaggcgcc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgcccac	atcaaggcac	cctgcctcgc	gcgtttcggt	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccgga	ctgtcacagc	ttgtctgtaa	gcggtacccg	3900
ggagcagaca	agcccgtcag	ggcgcgtcag	cgggtgttgg	cgggtgtcgg	ggcgacagcca	3960
tgaccagctc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcac	aggcgctctt	ccgcttcttc	gctcactgac	tcgctgcgct	cggctcgttcg	4140
gctgcgccga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggcgcggttg	ctggcgtttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttcccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gaccctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgctttc	tcatagtcca	cgctgtagg	atctcagttc	4500
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccgttc	agcccagaccg	4560
ctgcgcctta	tccggtaaact	atcgtcttga	gtccaaaccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtagggc	gtgctacaga	4680
gttcttgaag	gattggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	4800
caccgctgg	agcgggtggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gattccttga	tcttttctac	gggggtctgac	gctcagtgga	acgaaaaactc	4920
acgttaagg	atlttggtca	tgcattctag	gtactaaaac	aattcatcca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gtcgcacata	5040
ctgttcttcc	ccctgatcga	cccgacgcag	cgggacgcag	aaggcaatgt	cataccactt	5100
gtccgcccctg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgccgtggga	aaagacaagt	tcctcttcgg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgcgcggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tccggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctatcc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatcacgc	tccgataatct	tttcagggtc	ttgttcatct	tcatactctt	ccgagcaaac	5460
gagcccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgcccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcataggtg	gtccctttat	accggtgtc	cgctattttt	aaatataggt	tttcattttc	5640
ttccaccagc	ttatatacct	tagcaggaga	tattccttcc	gtatctttta	cgcagcggtg	5700
tttttcgatac	agttttttca	attccggtga	ctttctcatt	ttagccattt	attatttctc	5760
tctctttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcaactgt	tccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaagt	tgccgtataa	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcagcaa	cgctctgtca	tcggtacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaac	6060
tctgcccctc	tacaacggct	ctcccgtga	cgcgcgtccc	gactgatggg	ctgctgtat	6120
cgagtgggtg	ttttgtgcgg	agctgcccgt	cggggagctg	ttggctggct	gggtggcagga	6180
tatattgtgg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaattc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	tttagaaattt	tattgataga	agttattttac	aaatacaaat	acatactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tccagtcataa	tctcggtgac	gggcaggacc	6480

ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aaccacgcgc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tccgggtcgtt	gggcagcccg	atgacagcga	ccaogctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcctgt	ccgctggtgg	6720
cggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccttccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgcg	gacggacgag	gtcgtccgtc	cactcctgcg	gttccctgcg	ctcggtagcg	6900
aagttgaccg	tgcctgtctc	gatgtagtgg	ttgacgatgg	tgcagaccgc	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgtcgttctg	ggctcatggt	agactcgaga	7020
gagatagatt	tgtagagaga	gactggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatat	agaggaaggt	cttcggaagg	atagtggtat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	ggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcctttat	cgcaatgatg	gcattttagt	gtgccacctt	ccctttctac	7320
tgtccttttg	atgaagtgac	agatagctgg	gcaatggaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggtcttct	gagactgtat	ctttgatatt	7440
cttggagttag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tccgcagagg	catcttgaa	gatagccttt	cctttatcgc	aatgatggca	7620
ttttagtggt	ccaccttcct	tttctactgt	ccttttgatg	aaagtacaga	tagctgggca	7680
atggaatcgg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtccttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgtctagacc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tccggctcgt	7980
atgtttgtgt	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cggatatacag	acatgataag	atacatgtat	8100
gagtttgagc	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaattttgt	8160
gatgctattg	ctttattttgt	aaccattata	agctgcaata	aacaagttgg	ggtgggcgaa	8220
gaactccagc	atagatcccc	cgcgctggag	gatcatccag	ccggcgtccc	ggaaaaagat	8280
tccgaagccc	aacctttcat	agaaggcgcg	ctgggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgtctc	tccgcccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggccaactc	8400
ccgccccccac	ggctgctcgc	cgatctcggt	catggccggc	ccggaggcgt	cccgggaagt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacacca	8520
ggccagggtg	ttgtccggca	ccacctggtc	ctggaccgcy	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acaccggcga	agtcgtcctc	cacgaagtcc	cgggagaacc	cgagccggtc	8640
gggtccagaac	tgcacccgtc	cggcgacgtg	gcccgcgggt	agcaccggaa	cggcactggt	8700
caacttggcc	atggatccag	atttcgctca	aattagatata	aaaaagcagg	cttcaatcct	8760
gcagggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaagggc	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgcccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaagggaagg	ggcacctaca	8940
aatgcacagc	ttgcgataaa	ggaaaaggcta	tcgttcaaga	tgcctctgcc	gacagtggtc	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	ttggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	tacacagtc	tcagaagacc	aaagggtcat	tgagactttt	caacaaaagg	9180
taatattcgg	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaagggtgg	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgcccag	agtgggtcca	aagatggacc	cccaccacg	aggagcatcg	9360
tggaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaa	ggatgacgca	caatcccact	atccttcgca	agaccttctc	ctatataagg	9480
aagttcattt	cattttggaga	ggacacgctg	aaatcaccag	tctctctcta	caaactctac	9540
tctctcgagc	tttcgcagat	ccgggggggc	aatgagatat	gaaaaagcct	gaactcacgc	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctcogac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggaggggcg	ggatatgtcc	9720
tgcgggtaaa	tagctgogcc	gatgggttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcgcccg	gctcccgatt	cgggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgccgt	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgctgt	tctacaaccg	gtcgcggagg	ctatggatgc	gatcgtgcgc	gccgatctta	9960
gccagacgag	cgggttcggc	ccattcggac	cgcaagggaat	cgggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggacg	10080
acaccgtcag	tgcgtccgtc	gcgcagggtc	tcgatgagct	gatgcttttg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcgggtc	atgacttggg	gagaggcgat	gttcggggat	tcccaatacg	10260
aggtcgccaa	catcttcttc	tggaggccgt	ggttggcttg	tatggagcag	cagacgcgct	10320
acttcgagcg	gagggcatccg	gagcttgacg	gatcgccacg	actccggggc	tatatgctcc	10380
gcatttggct	tggaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440
gggcccaggg	tcgatgcgac	gcaatcgtcc	gatccggagc	cgggactgtc	gggctacac	10500

aaatcgcccg	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaaccg	acgccccagc	actcgtccga	gggcaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaaata	cttctatcaa	taaaatttct	aattcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcggttaa	ttcagatcaa	gcttgacctg	10860
gaatatcgcg	agtaaaactga	aaatcacgga	aaatgagaaa	tacacacttt	aggacgtgaa	10920
atatggcgag	gaaaaactgaa	aaagggtggaa	aatttagaaa	tgtccactgt	aggacgtgga	10980
atatggcaag	aaaactgaaa	atcatggaaa	atgagaaaaca	tccacttgac	gacttgaaaa	11040
atgacgaaat	cactaaaaaaa	cgtgaaaaaat	gagaaaatgca	cactgaagga	ctccgcggga	11100
attcgattgt	gctagccaat	gtttaacaag	atgtcaagca	caatgaatgt	tggtggttgg	11160
tggtcggtgg	tgggcggtggt	ggaaaattgc	ggtggttcga	gcggtagtga	tcggcgatgg	11220
ttggtgtttg	cagcggtggt	tgatagcgga	atcacttatg	gtggttggtca	caatggaggt	11280
gcgtcatggt	tattggtggt	tggtcatcta	tatatTTTTA	taataatatt	aagtatttta	11340
cctatttttt	acatatTTTT	tattaaattt	atgcattggt	tgtattttta	aatagttttt	11400
atcgacttgg	ttttataaaaa	tattttatta	ttttatgtgt	tatatattta	cttgatgtat	11460
tggaattttt	ctccattggt	ttttctatat	ttataataat	tttcttattt	ttttttgttt	11520
tattatgtat	tttttcgttt	tataataaat	atTTtattaaa	aaaaatatta	tttttgtaaa	11580
atatatcatt	tacaatgttt	aaaagtcatt	tgtgaatata	ttagctaagt	tgtacttctt	11640
tttgtgcatt	tgagtgttga	catgtctatt	atgattctct	ggccaaaaca	tgctactccc	11700
tgtcacttgg	gttttttttt	ttaaagacata	atcactagt	attatatcta	gactgaaggc	11760
gggaaacgac	aatctgatca	tgagcggaga	attaagggag	tcacgttatg	acccccggcg	11820
atgacgcggg	acaagccggt	ttacgtttgg	aactgacaga	accgcaacgt	tgaaggagcc	11880
actcagccgc	gggtttctgg	gctaagcaca	gactaagcaca	tacgtcagaa	accattattg	11940
cgcgttcaaa	agtcgcctaa	ggtcactatc	agctagcaaa	tatttcttgt	caaaaaatgct	12000
ccactgacgt	tcataaaatt	cccctcggt	tccaattaga	gtctcatatt	cactctcaat	12060
ccaaataatc	tgcaccggat	ctcgagatcg	aatTcccgcg	gccgcgaatt	cactagtggg	12120
tccccgggta	cggtcagtc	cttatgttac	gtcctgtaga	aaacccaacc	cgtgaaatca	12180
aaaaactcga	cggcctgtgg	gcattcagtc	tggtatcgcy	aaactgtgga	attgagcagc	12240
gttggtggga	aagcgcggt	caagaaagcc	gggcaattgc	tgtgccaggg	agtTTtaacg	12300
atcagttcgc	cgatgcagat	atcgttaatt	atgtgggcaa	cgtctggtat	cagcgcgaa	12360
tctttatacc	gaaagggttg	gcaggccagc	gtatcgtgct	gcgtttcgat	gcggtcactc	12420
attacggcaa	agtgtgggtc	aataatcagg	aagtgtgga	gcatacgggc	ggctatacgc	12480
catttgaagg	cgatgtcacg	ccgtatgtta	ttgccgggaa	aagtgtacgt	atcacagttt	12540
gtgtgaacaa	cgaactgaac	tggcagacta	tcccgccggg	aatggtgatt	accgacgaaa	12600
acgggaagaa	aaagcagtc	tacttccatg	atTTctttaa	ctacgccggg	atccatcgca	12660
gcgtaatgct	ctacaccacg	ccgaacacct	gggtggacga	tatcacctgt	gtgacgcatg	12720
tcgcgcaaga	ctgtaaccac	gcgtctgttg	actggcaggt	gggtggccaat	gggtgatgtca	12780
gcgttgaact	gcgtgatgcy	gatcaacagg	tggttgcaac	tggaacaaggc	accagcgagg	12840
ctttgcaagt	ggtgaatccg	cacctctggc	aatccgggtga	aggttatctc	tatgaactgt	12900
acgttcagag	caaaagccag	acagagtgtg	atatctaccc	gctgcgcgtc	ggcatccggt	12960
cagtggcagt	gaagggcgaa	cagttcctga	tcaaccacaa	accgttctac	tttactggct	13020
ttggccgtca	tgaagatgcy	gattttgcgcg	gcaaaaggatt	cgataacgtg	ctgatgggtg	13080
acgatcacgc	athtaaggac	tggtttgggg	ccaactccta	ccgtacctcg	cattaccctt	13140
acgctgaaga	gatgctcgac	tgggcagatg	aacatggcat	cgtgggtgatt	gatgaaactg	13200
cagctgtcgg	ctttaacctc	tctttaggca	ttggtttcga	agcgggcaac	aagccgaaag	13260
aactgtacag	cgaagaggca	gtcaacgggg	aaactcagca	ggcgcaacta	caggcgatta	13320
aagagctgat	agcgcgtagc	aaaaaccacc	caagcgtggt	gatgtggagt	attgccaaacg	13380
aaccggatac	ccgtccgcaa	ggtgcacggg	aatatttcgc	gccactggcg	gaagcaacgc	13440
gtaaactcga	tccgacgcgt	ccgatcacct	gcgtcaatgt	aatgttctgc	gacgtcacca	13500
ccgataccat	cagcgatctc	tttgatgtgc	tgtgcctgaa	ccgttattac	ggttggatatg	13560
tccaaagcgg	cgatttggaa	acggcagaga	aggtaactgga	aaaagaactt	ctggcctggc	13620
aggagaaact	gcatacggcg	attatcatca	ccgaatacgg	cgtggatacg	ttagccgggg	13680
tgcactcaat	gtacaccgac	atgtggagtg	aagagtatca	gtgtgcatgg	ctggatatgt	13740
atcacccgct	ctttgatcgc	gtcagcgccg	tcgtcgtgta	acaggtatgg	aatTTcgccg	13800
atTTtgcgac	ctcgcaaggc	atattgcgcg	ttggcggtaa	caagaagggg	atcttcaacc	13860
gcgaccgcaa	accgaagtgc	gcggcttttc	tgctgcaaaa	acgctggact	ggcatgaact	13920
tcggtgaaaa	agcccgagc	ggaggcaaac	aatgaatcaa	caactctcct	ggcgaccat	13980
cgtcggtac	agcctcgga	attgcgtagc	gagctcgaat	ttccccgatc	gttcaaacat	14040
ttggcaataa	agtttcttaa	gattgaatcc	tgttgcgggt	cttgcatgta	ttatcatata	14100
atTTctgttg	aattacgtta	agcatgtaat	aattaaacatg	taatgcatga	cgttatTTat	14160
gagatgggtt	tttatgatta	gagtcggca	attatacatt	taataacgca	tagaaaacaa	14220
aatatagcgc	gcaaactagg	ataaattatc	gcgcgcgggtg	tcatctatgt	tactagatcg	14280
ggaattcgat	atcaagcttg	gcactgggcy	tcgttttaca	acgtcgtgac	tgggaaaaacc	14340
ctggcggtac	ccaacttaac	cgccctgcag	cacatccccc	tttcgccagc	tgccgttaata	14400
gcgaagaggc	ccgcaccgat	cgcccttccc	aacagttgcy	cagcctgaat	ggcgaatgct	14460
agagcagctt	gagcttggat	cagattgtcg	tttcccgctt	tcagtttaaa	ctatcagttg	14520

-71-

ttgacaggat atattggcgg gtaaacctaa gagaaaagag cgtttattag aataacggat 14580
 attttaaagg gcgtgaaaag gtttatccgt tcgtccattt gtatgtg 14627

<210> 110
 <211> 9080
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18attBZeo (6XHS4) 2eGFP Plasmid

<400> 110
 cagttgccgg ccgggtcgcg cagggcggaac tcccgccccc acggetgctc gccgatctcg 60
 gtcattggcgg gcccgggagg gtcccgggaag ttcgtggaca cgacctccga ccactcggcg 120
 tacagctcgt ccaggcccgcg caccacacac caggccaggg tgttgccgg caccacctgg 180
 tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtctgct 240
 tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
 tcgctcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttctgct 360
 caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatata gaattcctgc 420
 agcccccgcg atccgctcac ggggacagcc ccccccaaaa gccccaggg atgtaattac 480
 gtccctcccc cgctaggggg cagcagcgag ccgccccggg ctccgctccg gtccggcgct 540
 ccccccgcat ccccgagccg gcagcgtcg gggacagccc gggcaccggg aaggtggcac 600
 gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
 acggggcgcc ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 720
 acgtccctcc ccgctaggg ggcagcagcg agccgcccgg ggctccgctc cgggtccggcg 780
 ctccccccgc atccccgagc cggcagcgtg cggggacagc ccggggcacc ggaaggtggc 840
 acgggatcgc tttcctctga acgcttctcg ctgctctttg agcctgcaga cactggggg 900
 atacggggcc cgggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960
 ttacgtccct cccccgctag ggggcagcag cgagccgccc ggggctccgc tccgggtccg 1020
 cgctcccccc gcattccccg gccggcagcg tcgggggaca gcccgggcac ggggaaggtg 1080
 gcacgggata gctttcctct gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
 ggatacgggg ccgcggtacc gctcacgggg acagcccccc cccaaagccc ccagggatgt 1200
 aattacgtcc ctcccccgct agggggcagc agcgagccgc ccgggggctcc gctccggtcc 1260
 ggcgctcccc ccgcatcccc gagccggcag cgtgcgggga cagcccgggc aggggaaggt 1320
 tggcacggga tcgctttcct ctgaacgctt ctgctgctc tttgagcctg cagacacctg 1380
 ggggatacgg ggccgcggtat ccgctcacgg ggacagcccc cccccaaagc cccagggat 1440
 gtaattacgt cctccccccg ctagggggca gcagcgagcc gcccggggct ccgctccggt 1500
 ccggcgctcc ccccgcatcc ccgagccggc agcgtgcggg gacagcccg gcacggggaa 1560
 ggtggcacgg gatcgctttc ctctgaacgc ttctcgctgc tctttgagcc tgcagacacc 1620
 tgggggatac gggggcgcgg atccgctcac ggggacagcc ccccccaaaa gccccaggg 1680
 atgtaattac gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg 1740
 gtcccgctcc ccccgcatcc ccccgagccg gcagcgtcg gggacagccc gggcaccggg 1800
 aaggtggcac gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca 1860
 cctgggggat acggggcggg ggatccacta gttattaata gtaatcaatt acgggggtcat 1920
 tagttcatag ccacatatag gagttccggg ttacataact tacggtaaat ggcccgctg 1980
 gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa 2040
 cgccaatagg gactttccat tgacgtcaat ggggtgacta tttacggtaa actgcccact 2100
 tggcagtaca tcaagtgtat catatgccaa gtacgcccc tattgacgtc aatgacggt 2160
 aatggcccgcc ctggcattat gccagttaca tgaccttatg ggactttcct acttggcagt 2220
 acatctacgt attagtcata gctattacca tgggtcgagg tgagccccc gttctgcttc 2280
 actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat tttttaatta 2340
 ttttgtgcag cgatgggggg gggggggggg ggggcgcgcg ccaggcgggg cggggcgggg 2400
 cgagggggcg ggcggggcga ggcggagagg tgcggcgga gccaatcaga gcggcgcgct 2460
 ccgaaagtct ccttttatgg cgaggcgggc gcggcgggc ccctataaaa agcgaagcgc 2520
 gcggcgggcg ggagtcgtg cgttgcttcc gcccgtgccc ccgctccgcg ccgctccg 2580
 ccgcccggcc cggtctcgac tgaccgcgtt actcccacag gtgagcgggc gggacggccc 2640
 ttctcctccg ggctgtaatt agcgttggg ttaatgacgg ctcgcttctt ttctgtggct 2700
 gcgtgaaagc cttaaagggc tccgggagg ccctttgtgc gggggggagc ggctcggggg 2760
 gtgctgtgct gtgtgtgtgc gtggggagcg ccgctgctgc ccgctgctgc ccggcggtg 2820
 tgagcgtgct gggcgcgggc ggggctttg tgcgtccgc gtgtgcgcga ggggagcgcg 2880
 gccggggggc gtgccccgcg gtgcgggggg gctgcgaggg gaacaaaggc tgcgtgcggg 2940
 gtgtgtgctg ggggggggta gcagggggtg tgggcgcggc ggtcgggctg taaccccc 3000
 ctgcaccccc ctccccgagt tgctgagcac gggccggctt cgggtgcggg gctccgtgcg 3060
 gggcgtggcg cggggctcgc cgtgccgggc ggggggtggc ggcaggtggg ggtgccgggc 3120
 gggggcgggc cgcctcgggc cggggaggcg gccattgcct tttatggtaa cctgcgaga 3180
 ccggcggtg gggcgaggg acttcccttg tcccaaatct ggcggagccg aaatctggga ggcgcgcgcg 3240
 gggcgaggg acttcccttg tcccaaatct ggcggagccg aaatctggga ggcgcgcgcg 3300

-72-

```

cacccccctct agcggggcgcg ggcgaagcggg tgcggcgccg gcaggaagga aatgggcggg 3360
gagggccttc gtgcgtcgcc gccgccgccgt ccccttctcc atctccagcc tcgggggctgc 3420
cgagggggga cgtctgcctt cggggggggac gggggttcggc tctggcgctg 3480
tgaccggcggt ctctagagcc tctgctaacc atgttcatgc cttcttcttt ttccctacagc 3540
tcctgggcaa cgtgctggtt gttgtgctgt ctcatcattt tggcaaagaa ttccgccacca 3600
tgggtgagcaa gggcgaggag ggttgcaccg ggggtggtgcc catcctggtc gagctggagc 3660
gcgacgtaaa cggccacaag ttcagcgtgt ccggcgaggg cgaggcgcat gccacctacg 3720
gcaagctgac cctgaagtcc atctgcacca ccggcaagct gcccggtgcc tggcccaacc 3780
tcgtgaccac cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc 3840
agcacgactt cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct 3900
tcaaggacga cggcaactac ggcaccccg cagagtgaa gttcgagggc gacacctgg 3960
tgaaccgcat cgagctgaag ggcacgact tcaaggagga cggcaacatc ctggggcaca 4020
agctggagta caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg 4080
gcattcaagg gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg 4140
accactacca gcagaacacc cccatcgggc acggccccgt gctgctgccc gacaaccact 4200
acctgagcac ccagtccgcc ctgagcaaaag accccaacga gaagcgcgat cacatggtcc 4260
tgctggagtt cgtgaccgcc gccgggataca ctctcgcat ggacgagctg tacaagtaag 4320
aattcactcc cttgggtcag gctgcctatc agaagtggtt ggctggtgtg gccaatgccc 4380
tggctcacia ataccactga gatctttttt cctctgccaa aaattatggg gacatcatga 4440
agccccctga gcatctgact tctggctaata aaggaattt tattttcatt gcaatagtgt 4500
gttggaaattt tttgtgtct tcaactcgaa ggacatatgg gagggcaaat catttaaac 4560
atcagaatga gtatttggtt tagagtttgg caacatatgc catatgctgg ctgccatgaa 4620
caaagggtggc tataaagagg tcatcagtat atgaaacagc cccctgctgt ccattcctta 4680
ttccatagaa aagccttgac ttgaggttag atttttttta tattttgttt tgtgttattt 4740
ttttctttaa catccctaaa agctgtccct cttctcttat tagccagatt ttctctctc 4800
tcctgactac tcccagtcac agctgtccct cttctcttat gaagatccct cgacctgcag 4860
cccaagcttg catgcctgca ggtgcactct agtggatccc ccgccccgta tccccagggt 4920
gtctgcaggc tcaaagagca gcgagaagcg ttcagaggaa agcgatccc tgccaccttc 4980
cccgtgcccc ggctgtcccc gcacgctgcc ggctcgggga tgcgggggga gcgccggacc 5040
ggagcgagcg cccggggcg cctgtgctgc cccctagcgg gggaggagcg taattacatc 5100
cctgggggct ttgggggggg gctgtccccg tgagcgatc cgcgcccccg tatccccag 5160
gtgtctgcag gctcaaagag cagcgagaag cgttcagagg cgttcagagg aaagcgatcc cgtgccacct 5220
tccccgtgcc cgggctgtcc ccgcacgctg ccggctcggg gatgcggggg gagcgccgga 5280
ccggagcgga gccccggggc gctcgtgct gctcgtgccc gccccctagc gggggaggga cgtaatata 5340
tccctggggg ctttgggggg gggtctgccc cgtgagcgga tccgcgcccc cgtatcccc 5400
aggtgtctgc aggtctaaag agcagcgaga agcgttcaga ggaaagcgat cccgtgccac 5460
cttccccgtg cccgggctgt cccgcacgc cccgcacgc tgcgggctcg gggatgcccc gggagcgccg 5520
gaccggagcg gagccccgg ggctcgtctg cgtcccccta gcgggggagg gacgttaatta 5580
catccctggg ggcttggggg gggggctgtc cccgtgagcg gatccgcccc cccgtatccc 5640
ccaggtgtct gcaggctcaa agagcagcga gaagcgttca gaggaaagcg atccccgtgc 5700
accttccccg tgccccgggt gtcgccgcac gctgccggct cggggatgct gggggagcgc 5760
cggaccggag cggagccccg ggcggtctgc tgcgtcccc tagcggggga gggagcgtat 5820
tacatccctg ggggctttgg ggggggctg tccccgtgag cggatccgcg gccccgtatc 5880
ccccaggtgt ctgcaggctc aaagagcagc gagaagcgtt cagaggaaag cgatcccgtg 5940
ccacctccc cgtccccgg cgtgtcccc cgtgtcccc ctcggggatg cgggggagc 6000
gccggaccgg agcggagccc ggggcggctc gctgtgccc cctagcgggg gagggacgta 6060
attacatccc tgggggcttt tggggggggg tgtccccgt agcggtatcc cggccccgta 6120
tccccagggt gcttgcaggc tcaaagagca gcgagaagcg ttcagaggaa agcgatccc 6180
tgccaccttc cccgtgcccc ggctgtcccc gcacgctgcc ggctcgggga tgcgggggga 6240
gcgccggacc ggagcggagc cccggcgggc tgcgtgctgc cccctagcgg gggaggagcg 6300
taattacatc cctggggggt tggggggggg gctgtcccc tgagcggtat cgcggggctg 6360
caggaattcg taatcatggt catagctgtt tccgtgtgta aattgttatc cgctcacaat 6420
tccacacaac atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagttag 6480
ctaactcaca ttaattgctg gctgctcact gcccgttttc cagtcgggaa acctgtcgtg 6540
ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttggcgctc 6600
ttccgcttcc tcgctcactg actcgctgct ctcggtcggt cggctgcggc gagcggtatc 6660
agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa 6720
catgtgagca aaaggccagc aaaaggccag ggcgtttccc cctggaagct tgctggcggt 6780
tttccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgctcaa gtccagaggtg 6840
gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgcg 6900
ctctcctggt cccgacctgc cgcttaccgg atacctgtcc gctttctcc cttcggaag 6960
cgtggcgctt tctcatagct cacgctgtag gtatctcagt tccgtgtagg tcgttcgctc 7020
caagctggcg tgtgtgcagc aacccccgt tcagcccgac cgctgcgcct tatccggtaa 7080
ctatcgtctt gagtccaacc cggtaaagca cgacttatcg ccactggcag cagccactgg 7140
taacaggatt agcagagcga cgtatgtagg cgggtctaca gagtcttga agtggtggcc 7200
taactacggc tacactagaa ggcagttatt gctctgctga agccagttac 7260
cttcggaaaa agagtttgta gctcttgatc cggcaacaa accaccgctg gtagcggtgg 7320

```


-73-

```

tttttttgtt  tgcaagcagc  agattacgcg  cagaaaaaaa  ggatctcaag  aagatccttt  7380
gatcttttct  acgggggtctg  acgctcagtg  gaacgaaaaa  tcacgttaag  ggatttttgg  7440
catgagatta  tcaaaaagga  tcttcaccta  gatcctttta  aattaaaaat  gaagttttta  7500
atcaatctaa  agtatatatg  agtaaacttg  gtctgacagt  taccaatgct  taatcagtga  7560
ggcacctatc  tcagcgatct  gtctatttct  ttcatccata  gttgcctgac  tccccgtcgt  7620
gtagataact  acgatacggg  agggccttacc  atctggcccc  agtgctgcaa  tgataccgcg  7680
agacccacgc  tcacgggctc  cagattttatc  agcaataaac  cagccagccg  gaagggccga  7740
gcgacagaagt  ggtcctgcaa  ctttatccgc  ctccatccag  totattaatt  gttgccggga  7800
agctagagta  agtagttcgc  cagttaatag  tttgcgcaac  gttgttgcca  ttgctacagg  7860
catcgtggtg  tcacgctcgt  cgtttggtat  ggcttcattc  agctccgggt  cccaacgata  7920
aaggcgagtt  acatgatccc  ccatgttggt  caaaaaagcg  gttagctcct  tcggtcctcc  7980
gatcgttgct  agaagtaagt  tggccgcagt  gttatcactc  atggttatgg  cagcactgca  8040
taattctctt  actgtcatgc  catccgtaag  atgcttttct  gtgactgggt  agtactcaac  8100
caagtcatct  tgagaatagt  gtatgcccgc  accgagttgc  tcttgcccgg  cgtcaatacg  8160
ggataatacc  gcgccacata  gcagaacttt  aaaagtgcct  atcattggaa  aacgttcttc  8220
ggggcgaaaa  ctctcaagga  tcttaccgct  gttgagatcc  agttcgatgt  aaccactcgt  8280
tgaccccaac  tgatcttcag  catcttttac  tttcaccagc  gtttctgggt  gagcaaaaaa  8340
aggaaggcaa  aatgccgcaa  aaaagggaat  aagggcgaca  cggaaatgtt  gaatactcat  8400
actcttctct  tttcaatatt  attgaagcat  ttatcagggt  tattgtctca  tgagcggata  8460
catatttgaa  tgtatttaga  aaaataaaca  aatagggggt  ccgcgcacat  tccccgaaa  8520
agtgccacct  gcagtagtta  acaaaaaaaa  gcccgccgaa  gcgggcttta  ttaccaagcg  8580
aagcgccatt  cgccattcag  gctgcgcaac  tgttggaag  ggcgatcggt  gcgggcctct  8640
tcgctattac  gccagctggc  gaaaggggga  tgtgctgcaa  ggcgattaa  ttgggttaac  8700
ccagggtttt  cccagtcacg  acgttgtaaa  acgacggcca  gtccgtaata  cgactcactt  8760
aaggccttga  ctagagggtc  gacgggtatac  agacatgata  agatacattg  atgagtttgg  8820
acaaaccaca  actagaatgc  agtgaaaaaa  atgctttatt  tgtgaaattt  gtgatgctat  8880
tgctttattt  gtaaccatta  taagctgcaa  taaacaagtt  ggggtgggcg  aagaactoca  8940
ccatgagatc  cccgcgtgg  aggatcatcc  agccggcgct  ccgaaaaacg  attccgaagc  9000
ccaacctttc  atagaaggcg  gcggtggaat  cgaaatctcg  tagcacgtgt  cagtctgtct  9060
cctcgccac  gaagtgcacg  9080

```

<210> 111

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA10 Plasmid

<400> 111

```

gttaactacg  tcaggtggca  cttttcgggg  aaatgtgctg  ggaaccctta  tttgtttatt  60
tttctaata  cattcaaata  tgtatccgct  catgagacaa  taaccctgat  aaatgcttca  120
ataatattga  aaaaggaaga  gtatgagtat  tcaacatttc  cgtgtcgccc  ttattccctt  180
ttttgcccga  ttttgccctc  ctgttttttg  tcaccagaa  acgctgggtg  aagtaaaaga  240
tgctgaagat  cagttgggtg  cacgagtggt  ttacatcgaa  ctggatctca  acagcggtaa  300
gatccttgag  agttttcgcc  ccgaagaacg  ttctccaatg  atgagcactt  ttaaagtctt  360
gctatgtggc  gcgggtattat  ccggtgttga  cgccgggcaa  gagcaactcg  gtcccgcat  420
acactattct  cagaatgact  tggttgagta  ctaccagtc  acagaaaagc  atcttacgga  480
tggcatgaca  gtaagagaat  tatgcagtg  tgccataacc  atgagtata  aactgaggc  540
caacttactt  ctgacaacga  tcggaggacc  gaaggagcta  accgcttttt  tgcacaacat  600
gggggatcat  gtaactcgcc  ttgatcggtg  ggaaccggag  ctgaatgaag  ccataccaaa  660
cgacgagcgt  gacaccacga  tgcctgtagc  aatggcaaca  acgttgcgca  aactattaac  720
tggcgaaact  cttactctag  cttcccgga  acaattaata  gactggatgg  aggcggataa  780
agttgcagga  ccacttctgc  gctcgccct  tccggctggc  tgggtttatt  ctgataaatc  840
tggagccgg  gagcgtgggt  ctgcgggtat  catgagca  ctggggccag  atggtaagcc  900
ctcccgatc  gtgattatct  acacgacggg  gagtcaggca  actatggatg  aacgaaatag  960
acagatcgct  gagataggtg  cctcactgat  taagcattgg  taactgtcag  accaagttta  1020
ctcatatata  ctttagattg  atttaccocg  gttgataatc  agaaaagccc  caaaaacagg  1080
aagattgtat  aagcaaatat  ttaaattgta  aacgttaata  ttttgtaaaa  attcgcgtta  1140
aatttttgtt  aaatcagctc  attttttaac  caataggccg  aaatcggcaa  aatcccttat  1200
aaatcaaaag  aatagcccga  gataggggtg  agtggtgttc  cagtttgcaa  caagagtcca  1260
ctattaaaga  acgtggactc  caacgtcaaa  gggcgaaaaa  ccgtctatca  gggcgatggc  1320
ccactacgtg  aacctacc  caaatcaagt  tttttggggt  cgaggtgccg  taaagcacta  1380
aatcggaacc  ctaaaaggag  ccccgattt  agagcttgac  ggggaaagcg  aacgtggcga  1440
gaaagggaag  gaagaaagcg  aaaggagcgg  gcgctagggc  gctggcaagt  gtagcggta  1500
cgctgcgct  aaccaccaca  cccgcgcgc  ttaaatgcgc  gctacagggc  gcgtaaaagg  1560
atctaggtga  agatcctttt  tgataatctc  atgacaaaaa  toctttaacg  tgagttttcg  1620

```

-74-

```

ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccttttttt 1680
ctgcgcgtaa tctgctgctt gcaaacacaaa aaaccaccgc taccagcggg ggtttggttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 1800
ccaaatactg ttcttctagt gtagcccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgcctacat acctcgctct gctaatoctg ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta cggggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 1980
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagagaaa ggcgagacagg 2100
tatccggtaa gggcgagggt cggaacagga gagcgcacga gggagcttcc agggggaaac 2160
gcctggtaac tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg 2220
tgatgtcgt caggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2280
ttcctggcct tttgctggcc ttttgcctac atgtaatgtg agttagctca ctcattaggc 2340
accccaggct ttacacttta tgcttccggc tegtatgttg tgtggaattg tgagcggata 2400
acaatttcac acaggaacaa gctatgacca tgattacgcc aagctacgta atacgactca 2460
ctagtggggc ccgtgcaatt gaagccggct ggcgccaagc ttctctgcag gatagaagcc 2520
tgctttttta tactaacttg agcgaatctt ggatcaccat gaaaacattt aacattttctc 2580
aacaagatct agaattagta gaagtagcga cagagaagat tacaatgctt tatgaggata 2640
ataaacatca tgtgggagcg gcaattcgta cgaaaacagg agaaatcatt tcggcagtac 2700
atattgaagc gtatatagga cgagtaactg tttgtgcaga agccattgcg attggtagt 2760
cagtttcgaa tggacaaaag gattttgaca cgattgtagc tggtagacac cctattctg 2820
acgaagttaga tagaagtatt cagtggttaa gtccttggg tatgtgtagg gagttgattt 2880
cagactatgc accagattgt tttgtgttaa tagaaatgaa tggcaagtta gtcaaaacta 2940
cgattgaaga actcattcca ctcaaatata cccgaaatta aaagttttac cataccaagc 3000
ttggctgctg cctgaggctg gacgacctcg cggagttcta ccggcagtg ccaatccgtc 3060
gcatccagga aaccagcagc ggctatccgc gcatccatgc ccccgaaactg caggagtggg 3120
gaggcacgat ggccgctttg gtccggatct ttgtgaagga accttacttc tgtggtgtga 3180
cataattgga caaactacct ctgattctaa aaagctctaa ggtaaatata aaatttttaa 3240
gtgtataatg tgttaaaacta ttgtttgtgt cctttaatga attttagatt ggaaaacctg 3300
gaactgatga atgggagcag tgggtggaat cctgtgactc tcaacattct ttttgcctag 3360
aagaaatgcc atctagtgat gatgaggcta ctgctgactc agaattgcta agttttttga 3420
aaaagaagag aaaggtagaa gacccaagg actttccttc agatttttga 3480
gtcatgctgt gtttagtaat agaactcttg cttgctttgc tatttacacc acaaaggaaa 3540
aagctgcact gctatacaag aaaatgatgg aaaaatatct tgtaaccttt ataagtaggc 3600
ataacagtta taatcataac atactgtttt tcttactcc acacaggcat agagtgtctg 3660
ctattataaa ctatgctcaa aaattgtgta cctttagctt ttttaattgt aaaggggtta 3720
ataaggaata tttgatgtat agtgccctga ctagagatca taatcagcca taccacattt 3780
gtagaggttt tacttgcttt aaaaaacctc ccactacctc ccctgaacct gaaacataaa 3840
atgaatgcaa ttgttgttgt taactgtgtt taatcgactt ataatgggta caaataaagc 3900
aatagcatca caaatttcac aaataaagat ccacgaattc gctagcttcg gccgtgacgc 3960
gtctccggat gtacaggcat gcgtcgaccc tctagtcaag gccttaagtg agtcgtatta 4020
cggactggcc gtgcgttttac aacgctcgtg ctgggaaaac cctggcggtta cccaacttaa 4080
tcgccttgca ccatcctccc ctttcgccag ctggcgtaat agcgaagagg cccgcaccga 4140
tcgcccttcc caacagttgc gcagcctgaa tggcgaaatg cgcttcgctt ggtaataaag 4200
ccgccttcgg cgggcttttt ttt 4223

```

<210> 112

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCX-LamIntR Plasmid

<400> 112

```

gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccatatat ggagttccgc gttacataac ttacgggtaaa tggcccgctt ggctgaccgc 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtac 240
atcaagtgtg tcatatgcc a gtacgccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360
tattagtcat cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctcccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
gcgatggggg cggggggggg gggggcgcg cccaggcggg gccggggcgg gcgaggggcy 540
ggggcgggcg agggcgagag gtgcggcggc agccaatcag agcggcgcg gcgaggggcy 600
tccttttatg gcgagggcgg cgcgttgctt gcccataaaa aagcgaagcg cgcggcgggc 660
gggagtcgt cgtttgcctt ccccgctcgc gcgcctcgc gccggcgcc cccggcgcc 720
ccggctctga ctgaccgcgt tactccaca ggtgagcggg cgggacggcc cttctcctcc 780

```


-75-

```

gggctgtaat tagcgcttgg tttaatgacg gctcggtttct tttctgtggc tgcgtgaaag 840
ccttaaaggg ctccgggagg gccctttgtg cggggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagg gccgcgtgcg ccggcgggct gtgagcgctg 960
cgggcggggc ggggggcttt gtgcgctccg cgtgtgcgcg aggggagcgc ggccgggggc 1020
ggtgccccgc ggtgcggggg ggctgcgagg ggaacaaagg ctgctgctgcg ggtgtgtgcg 1080
tggggggggtg agcagggggg gtggggcgcg cggtcgggct tccgggtgcg ggtgcacccc 1140
cctccccgag ttgctgagca cggcccggtc cgggtgtgcg ggctccgtgc ggggctgtgc 1200
gcggggctcg ccgtgccggg cgggggggtg cggcagggtg gggtgccggg cggggcgggg 1260
ccgcctcggg ccggggaggg ctccgggggag gggcgcgggc gccccggagc gccggcggtc 1320
gtcgaggcgc gggagccgc agccattgcc ttttatggta atcgctgcgag agggcgcgag 1380
gacttccctt gtcccaaate tggcggagcc gaaatctggg aggcgcggcc gcacccccctc 1440
tagcggggcg cggcgagagc gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt 1500
cgtgcgtcgc acggctgcct tcggggggga cggggcgagg ctcggggctg ctggggctg cgcaggggg 1560
acggctgcct gctctagagc ctctgctaac catgttcag cttctgttcggt gtagccggcg 1620
acgtgctggt tgttgtgtg tctcatcatt ttggcaaaga attcatggga agaaggcgaa 1680
gtcatgagcg cgggattta cccctaacc tttatataag aaacaatgga tattactgct 1740
acagggaccc aaggagcggg aaagagttt gattaggcag agacaggcga atcgcaatca 1800
ctgaagctat acaggccaac attgagttat tttcaggaca caaacacaag cctctgacag 1860
cgagaatcaa cagtgataat tccgttacgt tacattcatg gcttgatcgc taccgaaaaa 1920
tctggccag cagaggaatc gatgctccac cactcataaa ttacatgagc aaaattaaag 2040
caataaggag gggctcgcct gacgagggca aggcggcgct caccacaaaa gaaattgcgg 2100
caatgctcaa tggatacata gaggcaatag ctgaaggcca tataacaaca atcatgcaa 2160
cactgagcga tgcattccga gggcaatag ggagatcaag agccaagtta accatgtcg 2220
ctgccactcg cgcagcaaaa tctagagtaa catgttgggt tatgacagc acttacggct 2280
tgaaaattta tcaagcagca gaatcatcac gatgttgggt gaagtggtct atggaactgg 2340
ctgttgttac cgggcaacga gttgggtgatt tatgcgaaat aattgccatc gatatcgtag 2400
atggatatct tttatgcgag caaagcaaaa atagcgtaaa tgataaatgc ccaacagcat 2460
tgcatattga tgctctcgga atatcaatga aggaacact tgataaatgc aaagagattc 2520
ttggcgggag aaccataatt gcatctactc gtcgcgaacc gctttcatcc ggcacagtat 2580
caaggatatt tatgcgcgca ttgtctgcaa ctaagtcgg cttcgaaagg gatccgcta 2640
cctttcacga gttgcgcagt tttctcggg gctcctatga gaagcagata agcgataagt 2700
ttgctcaaca tcttctcggg cataagtcgg acaccatggc atcacagtat cgtgatgaca 2760
gaggcagggg gtgggacaaa attgaaatc aataagaatt cactcctcag gtgcaggctg 2820
cctatcagaa ggtggtggct atgcccggc tacaataac cactgagatc 2880
tttttccctc tgccaaaaat tatggggaca tcatgaagcc ccttgagcat ctgacttctg 2940
gctaataaag gaaatttatt ttcatgtcaa tagtgtgttg gaattttttg tgtctctcac 3000
tcggaaggac atatgggagg gcaaattcatt taataacatc gaattgagtat ttggtttaga 3060
gtttggcaac atatgccata tgctggctgc catgaacaaa ggtggctata aagaggtcat 3120
cagtatatga aacagccccc tgctgtccat tccttattcc atagaaaagc cttgacttga 3180
ggttagattt tttttatatt ttgtttttgt tttttttttt ctttaacatc cctaaaattt 3240
tccttacatg ttttactagc gatatttttc ctctctcct gactactccc agtcatagct 3300
gtccctcttc tcttatgaag atccctcgac ctgcagccca agcttggcgt aatcatggtc 3360
atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca taccagccgg 3420
aagcataaag ggtaaagcct ggggtgccta taactcacat taattgcgtt 3480
gcgtcactg ccgcctttcc agtcgggaaa cctgtcgtgc cagcggatcc gcatctcaat 3540
tagtcagcaa ccatagtccc gcccctaact ccgcccattc tccgcccagg 3600
tccgcccatt ctccgcccc cagaagtag tggctgacta tttatgcaga ggccgaggcc 3660
gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctaggcttt 3720
tgcaaaaagc taacttgttt attgcagctt ataatggtta caaataaagc aatagcatca 3780
caaatttcac aaataaagca tttttttcac tgcatctag ttgtggtttg tccaaactca 3840
tcaatgtatc ttatcatgtc tggatccgct gcattaatga atcggccaac gcgcggggag 3900
aggcggtttg cgtattgggc gctcttcgc ttctcgtc actgactcgc tgcgtcggg 3960
cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg tatccacaga 4020
atcaggggat aacgcaggaa agcaaaaagg cagcaaaaagg ccagcaaaaagg ccaggaaaccg 4080
taaaaaggcc gcgttgctgg cgtttttcca taggctcgcg cccctgacg agcatcaaaa 4140
aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcggt 4200
tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgcccgtta cgggatacct 4260
gtccgccttt ctcccttcgg gctccaagct gcttctcaa tgctcacgct gtaggtatct 4320
cagttcggtg taggtcgttc gctccaagct gggctgtgtg caccgaacccc ccgttcagcc 4380
cgaccgctgc gccttatccg gtaactatcg tcttgagtc aaccgggtaa gacacgactt 4440
atcgccactg gcagagcca ctggtaacag cgttagcaga gcgaggtatg taggcgggtg 4500
tacagagttc ttgaagtggg ggccctaacta cggctacact agaaggacag tatttggtat 4560
ctgcgctctg ctgaagccag ttaccttcgg ggtagctctt gatccggcaa 4620
acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta agtggaacga 4680
aaaaggatct caagaagatc ttttgatctt ttctacggg tctgacgctc cgcagagaaa 4740
aaactcacgt taagggattt tggctcatgag attatcaaaa aggatcttca cctagatcct 4800

```

-76-

tttaaatttaa	aaatgaagtt	ttaaatacaat	ctaaagtata	tatgagtaaa	cttgggtctga	4860
cagttaccac	tgcttaatac	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcac	4920
catagttgcc	tgactcccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	4980
ccccagtgct	gcaatgatac	cgcgagaccc	acgtccaccg	gctccagatt	tatcagcaat	5040
aaaccagcca	gccggaagg	cgcgagcgag	aagtggctct	gcaactttat	ccgcctccat	5100
ccagttctatt	aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	5160
caacgttggt	gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	5220
attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	cccccatgt	tgtgcaaaaa	5280
agcgggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtggtatc	5340
actcatgggt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	5400
ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	5460
ttgctcttgc	ccggcgctcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	5520
gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgtgtgtgag	5580
atccagttcg	atgtaacca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	5640
cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	5700
gacacggaaa	tggtgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	5760
gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	5820
ggttcgcgc	acatttcccc	gaaaagtgcc	acctg			5855

<210> 113

<211> 4346

<212> DNA

<213> Artificial Sequence

<220>

<223> pSV40-193AttpsensePur Plasmid

<400> 113

ccggtgccgc	caccatcccc	tgacccacgc	ccctgacccc	tcacaaggag	acgaccttcc	60
atgaccgagt	acaagcccac	gggtgcgcctc	gccaccgcgc	acgacgtccc	ccggggccgta	120
cgaccctcg	cgccgcggtt	cgccgactac	cccgcacgc	gccacaccgt	cgaccocggac	180
cgccacatcg	agcgggtcac	cgagctgcaa	gaactcttcc	tcacgcgcgt	cgggctcgac	240
atcggcaagg	tgtgggtcgc	ggacgacggc	gccgcggttg	cggctctggac	cacgcocggag	300
agcgtcgaag	cggggggcgg	gttcgccgag	atcggcccg	gcatggccga	gttgagcggg	360
tcccggctcg	ccgcgcagca	acagatggaa	ggcctcctgg	cgccgcaccg	gcccaggag	420
cccgcgtgg	tcctggccac	cgtcggcgctc	tcgcccagac	accagggcaa	gggtctgggg	480
agcgcgctcg	tgctcccccg	agtggaggcg	gccgagcgcg	ccgggggtgcc	cgccttcctg	540
gagacctccg	cgccccgcaa	cctccccctc	tacgagcggc	tcggcttcac	cgtcacccgc	600
gacgtcgagg	tgccccgaagg	accgcgcacc	tggtgcatga	cccgaagcc	cggtgcctga	660
cgccccgccc	acgacccgca	gcccgcgacc	gaaaggagcg	cacgacccca	tggtctccgac	720
cgaagccgac	ccggggcgcc	ccgcgcgacc	cgcaccccg	cccaggggcc	accgactcta	780
gaggatcata	atcagccata	ccacatttgt	agaggtttta	cttgccttaa	aaaacctccc	840
acacctcccc	ctgaacctga	aacataaaat	gaatgcaatt	gttggttgta	acttggttat	900
tgcagcttat	aattggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	960
tttttctactg	cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	atcatgtctg	1020
gatccgcgctc	ggatcccttaa	ttaagtctag	agtcgactgt	ttaaacctgc	aggcatgcaa	1080
gcttggcgta	atcatgggtca	tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacaaatc	1140
cacacaacat	acgagccgga	agcataaagt	gtaaaagcctg	gggtgcctaa	tgagtgaagt	1200
aactcacatt	aattgcgttg	cgctcactgc	ccgctttcca	gtcgggaaac	ctgtcgtgcc	1260
agctgcatta	atgaatcggc	caacgcgcgg	ggagaggcgg	tttgctgatt	gggcgctctt	1320
ccgcttcctc	gctcactgac	tcgctgcgct	cggctggtcg	gctgcggcga	gcggtatcag	1380
ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	1440
tgtagcaaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	ctggcggttt	1500
tccataggct	ccgccccctc	gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	1560
gaaacccgac	aggactataa	agataaccagg	cgtttccccc	tggaagctcc	ctcgtgcgct	1620
ctcctgttcc	gacctgccc	cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	1680
tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	1740
agctgggctg	tgtgcacgaa	cccccggttc	agcccgacgc	ctgcgcctta	tccggtaact	1800
atcgtcttga	gtccaaccgc	gtaagacacg	acttatcgcc	actggcagca	gccactggta	1860
acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tggtggccta	1920
actacggcta	cactagaagg	acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	1980
tcggaaaaag	agttggtagc	tcttgatccg	gcaaaacaa	caccgctggg	agcgggtggg	2040
ttttgttttg	caagcagcag	attacgcgca	gaaaaaaaag	atctcaagaa	gatcctttga	2100
tcttttctac	ggggtctgac	gctcagtgga	acgaaaaactc	acgttaaggg	attttgggtca	2160
tgagattatc	aaaaaggatc	ttcacctaga	ctctttttaa	ttaaaaatga	agtttttaaat	2220
caatctaagc	tatatatgag	taaaacttggt	tcgacagtta	ccaatgctta	atcagtgagg	2280
cacctatctc	agcgatctgt	ctatttcggt	catccatagt	tgccctgactc	cccgtcgtgt	2340

-77-

```

agataactac gatacgggag ggcttaccat ctggcccccag tgctgcaatg ataccgcgag 2400
accacgctc accgggtcca gatttatcag caataaacca gccagccgga agggccgagc 2460
gcagaagtgg tcctgcgaact ttatccgcct ccacccagtc tattaattgt tgccgggaag 2520
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tggtgccatt gctacaggca 2580
tcgtgggtgtc acgctcgtcg tttgggtatgg cttcattcag ctccggttcc caacgatcaa 2640
ggcgagttac atgatcccc atgtttgtgca aaaaagcggg tagctccttc ggtcctccga 2700
tcgtttgtcag aagtaagtgt gccgcagtgt tatcaactcat gggtatggca gcaactgcata 2760
attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 2820
agtcatcttg agaatagtgt atggggcgac cgagttgtct ttgcccggcg tcaatacggg 2880
ataataccgc gccacatagc agaactttaa aagtgtcat cattggaaaa cgttcttcgg 2940
ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 3000
caccacactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 3060
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 3120
tcttcctttt tcaatattat tgaagcattt atcaggggtta ttgtctcatg agcggatata 3180
tatttgaatg tatttagaaa aataacaaa taggggttcc gcgcacattt ccccgaaaag 3240
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 3300
tcacgaggcc ctttcgtctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 3360
agctcccggg gacggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 3420
aggcgcgctc agcgggtgtt ggcgggtgtc ggggctgggt taactatgcg gcatcagagc 3480
agattgtact gagagtgcac catatgcggg gtgaaatacc gcacagatgc gtaaggagaa 3540
aataccgcat caggcgccat tcgccattca gcttgcgcaa ctgttgggaa gggcgatcgg 3600
tgccggcctc ttcgctatta cgccagctgg cgaagggggg atgtgctgca aggcgattaa 3660
gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtgaattcg 3720
agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggcctccca gcaggcagaa 3780
gtatgcaaaag catgcatctc aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3840
cagcaggcag aagtatgcaa agcatgcac tcattagtc agcaaccata gtcccgcctc 3900
taactccgcc catcccgccc ctaactccgc ccagttccgc ccattctccg ccccatggct 3960
gactaatttt ttttatttat ggaggctcgg taccctcttg ggcctctgag ctattccaga 4020
agtagtgagg aggtttttt gttgattcat agtgactgca cgctaattgt ctgttacagg 4080
tactaataac catctaagta gttgattcat tatgttgtgt ttacagtat 4140
tatgtagtct gttttttatg caaaatctaa tttaatatat tgatatttat atcattttac 4200
gtttctcggt cagctttttt atactaagtt ggcattataa aaaagcattg cttatcaatt 4260
tggttcaacg aacaggtcac tatcagtcac aataaaatca ttatttgatt tcaattttgt 4320
cccactccct gcctctgggg ggcgcg 4346

```

<210> 114

<211> 3166

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo Plasmid

<400> 114

```

cagttgccgg cccgggtcgg caggggcgaac tcccgccccc acggtctgct gccgatctcg 60
gtcatggccg gcccgagggc gtcccggaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccg cccccacacc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtctgtc 240
tccacgaagt cccgggagaa ccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgcattg ctgcaggctg 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttctc 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctacatttaa ttgctgtgag ctactgccc 600
gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgctcgggg 660
agaggcggtt tgcgtattgg gcgctcttcc gcttccctcg tcaactgact gctgcgctcg 720
gtcgttcggc tgcggcgagc ggtatcagct cactcaaaag cggtaatacg gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggcagggaac 840
cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtca gaggtggcga aaccggacag gactataaag ataccaggcg 960
tttcccctcg gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac 1020
ctgctcggct ttctcccttc gggaaagcgt gcgctttctc atagctcacg ctgtaggatg 1080
ctcagctcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag 1140
cccgaccgct gcgccttata cgtcttgagt ccaaccgggt aagacacgac 1200
ttatcgccac tggcagcagc cactggtaac gagcgaggta tgtaggcggt 1260
gctacagagt tcttgaagtg gtggcctaac ctagaaggac agtatctggt 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

```

-78-

```

aaacaaacca ccgctggtag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tccttttgatc ttttctacgg ggtctgacgc tcagtggaaac 1500
gaaaactcac gtttaaggat tttggtcatg agattatcaa aaaggatctt cacttagatc 1560
cttttaaat aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 1620
gacagttacc aatgcttaat cagtggaggca cctatctcag cgatctgtct atttcgttca 1680
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaaccagc cagccggaag ggccgagcgc agaagtggc ctgcaacttt atccgcctcc 1860
atccagtcct ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg 1920
cgcaacgttg ttgcatggc tacaggcatc gtggtgtcac gctcgtcgtt tgggtatggct 1980
tcattcagct ccggttccca acgatcaagg cgagttacat gatccccc atgtgtgcaaa 2040
aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagtggc cgcagtgtta 2100
tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc 2160
ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 2220
agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa 2280
gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 2340
agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc 2400
accagcgttt ctgggtgagc aaaaacagga aggcataatg ccgcaaaaaa gggaataagg 2460
gcgacacgga aatgttgaa actcatactc ttctttttc aatattattg aagcatttat 2520
caggttatt gtctcatgag cggatacata ttgaaatgta tttagaaaaa taaacaaata 2580
ggggttcgcg gcacatttc ccgaaaagt ccactgacg tagttaacaa aaaaaagccc 2640
gccgaagcgg gctttattac caagcgaagc gccattcgcc attcaggctg cgcaactggt 2700
gggaagggcg atcgggtcgg gcctcttcgc tattacgcca gctggcgaaa gggggatgtg 2760
ctgcaaggcg attaagttgg gtaacgccag ggtttccca gtcacgacgt tgtaaaaacga 2820
cggccagtc gtaatacgac tcacttaagg ccttgactag agggtcgacg gtatacagac 2880
atgataagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc 2940
tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag ctgcaataaa 3000
caagttgggg tgggcgaaga actccagcat gagatccccg cgctggagga tcatccagcc 3060
ggcgtcccg gaaacgattc cgaagcccaa cctttcatag aaggcggcgg tgggaatcgaa 3120
atctcgtagc acgtgtcagt cctgctcctc ggccacgaag tgcacg 3166

```

<210> 115

<211> 7600

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo3'6XHS4eGFP Plasmid

<400> 115

```

cagttgccgg cggggtcgcg cagggcgaac tcccgcccc acggctgctc gccgatctcg 60
gtcattggccg gcccgaggc gtcccggaag ttctgtgaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccg caccacacac caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgtgatgaa cagggtcacg tcgtcccgga ccacaccggc gaagtcgtcc 240
tccacgaagt ccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatcta gttattaata 420
gtaatacaatt acgggtcat tagttcatag cccatataat gagttccgcg ttacataact 480
tacggtaaat ggccgcctg gctgaccgcc caacgacccc cgccattga cgtcaataat 540
gacgtatgtt cccatagtaa cgccaatagg gactttccat tgacgtcaat ggggtggacta 600
tttacggtaa actgcccact tggcagtaca tcaagtgtat catatgccaa gtacgcccc 660
tattgacgtc aatgacggt aatggcccg cttggcattat gccagtaga tgaccttatg 720
ggactttcct acttggcagt acatctacgt attagtcac gctattacca tgggtcgagg 780
tgagccccac gttctgctt actctcccc tctccccccc ctccccacc ccaattttgt 840
atttatttat ttttaatta tttgtgcag cgtggtgggc gggggggggg ggggcgcgcg 900
ccaggcgggg cggggcgggg cgagggggcg ggccgggaga ggccgagagg tgcggcgcca 960
gccaatcaga gcggcgcgct ccgaaagttt ccttttatgg cgaggcgggc gcggcgggcg 1020
ccctataaaa agcgaagcgc ccgcccggcg ggagtcgctg cgttgccctc gcccggtg 1080
ccgctccgcg ccgctcgcg ttctcctccg ggctgtaatt agcgtctggt ttaatgacgg 1140
gtgagcgggc gggacggccc gctgaaagc cttaaagggc tccgggaggg ccctttgtgc 1200
ctcgtttctt ttctgtggc gtgcgtgcgt gtgtgtgtgc ggggggagcg ccgcgctgg 1260
gggggggagc ggctcggggg tgagcgtgct gggcgcgggc cggggctttg tgcgctccgc 1320
cccgcgtgc cggcggtg gtgagcgtgc ggcggggggc gtgcccgcg gtcgggggg 1380
gtgtgcgcga ggggagcgc gtcgtgcgg gtgtgtgcgt ggggggggtg gcagggggtg 1440
gaacaaaggc tgccgtgcgg gtcgaccccc ctcgccgagt tgctgagcac ggcccggtt 1500
ggcgggctg gctccgtgcg gggcgtggcg cggggtcgc cgtgccgggc ggggggtggc 1560

```

-79-

ggcaggtggg	ggtgccgggc	ggggcggggc	cgctcgggc	cggggagggc	tgggggagg	1680
ggcgcgggcg	ccccggagcg	cgggcgggctg	tggggcgcg	gcgagccgca	gccattgcct	1740
tttatggtaa	tcgtgcgaga	gggcgaggg	acttcctttg	tcccaaatct	ggcgagccg	1800
aaatctggga	ggcgccgccc	cacccctct	agcgggcgcg	ggcgaagcgg	tggcgcccg	1860
gcaggaagga	aatgggcggg	gagggccttc	gtgctgcgc	gcgccgccc	ccccctctcc	1920
atctccagcc	tcggggctgc	cgaggggga	cggtgcctt	cgggggggac	ggggcgaggc	1980
gggggttcggc	ttctgctgtg	tgaccggcgg	ctctagagcc	tctgctaacc	atgttcatgc	2040
cttctctctt	ttcctacagc	tcctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	2100
tggcaaagaa	ttcgccacca	tggtgagcaa	ggcgaggag	ctgttcaccg	gggtggtgcc	2160
catcctggtc	gagctggacg	gcgacgtaaa	cgcccaacaag	ttcagcgtgt	ccggcgagg	2220
cgagggcgat	gccacctacg	gcaagctgac	cctgaagtct	atctgcacca	ccggcaagct	2280
gcccgtgccc	tggcccaccc	tcgtgaccac	cttgacctac	ggcgtgcagt	gcttcagccg	2340
ctaccccgac	cacatgaagc	agcacgactt	cttcaagtcc	gccatgcccg	aaggctacgt	2400
ccaggagcgc	accatcttct	tcaaggacga	cggaactac	aagaccgcg	ccgaggtgaa	2460
gttcgagggc	gacacctgg	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	2520
cggcaacatc	ctggggcaca	agctggagta	caactacaac	agccacaacg	tctatatcat	2580
ggccgacaag	cagaagaacg	gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	2640
cggcagcgtg	cagctcgccg	accactacca	gcagaacacc	cccatcgccg	acggccccgt	2700
gctgctgccc	gacaaccact	acctgagcac	ccagtcggcc	ctgagcaaa	accccaacga	2760
gaagcgcgat	cacatgggtc	tgctggagtt	ctgaccgccc	gccgggatca	ctctcgccat	2820
ggacgagcgt	tacaagtaag	aattcactcc	ttaggtgcag	gctgcctatc	agaaggtggt	2880
ggctgggtgtg	gccaatgccc	tggtccacaa	ataccactga	gatctttttc	cctctgcaa	2940
aaattatggg	gacatcatga	agccccctga	gcactgact	tctggcta	aaaggaaatt	3000
tattttcatt	gcaatagtgt	gttggaaatt	tttgtgtctc	tcactcgga	ggacatatgg	3060
gagggcaaat	catttaaaac	atcagaatga	gtatttggtt	tagagttag	caacatatgc	3120
catatgctgg	ctgccatgaa	caaaggtggc	tataaagagg	tcacagtat	atgaaacagc	3180
ccccctgctg	ccattcctta	ttccatagaa	aagccttgac	ttgaggttag	atttttttta	3240
tattttgttt	tgtgttattt	ttttctttaa	catccctaaa	attttcctta	catgttttac	3300
tagccagatt	tttctctctc	tcctgactac	tcacagtcac	agctgtccct	cttctcttat	3360
gaagatccct	cgacctgcag	cccaagcttg	catgctgca	ggtcgactct	agtggatccc	3420
ccgccccgta	tcccccaggt	gtctgcaggc	tcaaaagagca	gcgagaagcg	ttcagaggaa	3480
agcgatcccg	tgccaccttc	cccggtgccc	ggctgtcccc	gcacgctgcc	ggctcgggga	3540
tgccgggggga	gcgcccggacc	ggagcgggagc	cccgggcggc	tcgctgctgc	ccccctagcg	3600
gggagggggg	gaggggagc	ccggggggct	tcgggggggg	gctgtccccg	tgagcggtac	3660
cgcgggcccg	tatccccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cggttcagagg	3720
aaagcgatcc	cgtgccacct	tcctccgtg	cggtgtgtcc	ccgcacgctg	ccggctcggg	3780
gatcgggggg	gagcgccgga	ccggagcgga	gccccggggc	gctcgctgct	gccccctagc	3840
gggggagggga	cgtaattaca	tcctgggggg	ctttgggggg	gggtgtgtcc	cgtagagcga	3900
tcgcggggccc	cgtatcccc	aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	3960
ggaaagcgat	cccgtgccac	cttccccctg	cccgggtgt	ccccgcacgc	tgccggctcg	4020
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggtcgctg	ctgcccccta	4080
gcgggggagg	gagcgtaat	catccctggg	ggctttgggg	gggggctgtc	cccgtagcgc	4140
gatccgcggc	cccgtatccc	ccaggtgtct	gcaggtcaa	agagcagcga	gaagcgttca	4200
gaggaaagcg	atcccggtgc	accttccccg	tgcccggtgt	gtccccgcac	gctgcggct	4260
cggggatgcg	gggggagcgc	cggaacggag	cggaaccccg	ggcggtcgct	tgctgcccc	4320
tagcggggga	gggacgta	tacatccctg	ggggctttgg	gggggggctg	tcctcgtag	4380
cggatccgcg	gccccgtatc	ccccaggtgt	ctgcaggtc	aaagagcagc	gagaagcgtt	4440
cagaggaag	cgatcccggt	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgcggg	4500
ctcggggatg	cggggggagc	gcccggaccg	agcggaagcc	cgggcggtc	gctgctgccc	4560
cctagcgggg	gagggacgta	attacatccc	tggggggttt	gggggggggc	tgctccccgtg	4620
agcggatccg	cggccccgta	tcctccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	4680
ttcagaggaa	agcgatcccg	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	4740
ggctcgggga	tgccggggga	gcgcccggacc	ggagcgggagc	cccgggcggc	tcgctgctgc	4800
cccctagcgg	gggagggagc	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	4860
tgagcggatc	cgcggggctg	caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	4920
aattgttatc	cgctcacaat	tcacacaaac	atacgagccg	gaagcataaa	gtgtaaagcc	4980
tgggggtgct	aatgagttag	ctaactcaca	ttaattgctg	tgctgtcact	gcccgttttc	5040
cagtcgggaa	acctgtcgtg	ccagctgcac	taagtgaatc	gccaacgcgc	ggggagaggc	5100
ggtttgcgta	ttgggcgctc	ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcgtt	5160
cggtgctggc	gagcggatc	agctcactca	aaggcggtaa	tacggttatc	cacagaatca	5220
ggggataacg	caggaagaa	catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	5280
aaggcccgct	ttgggcgctt	ttcccaatag	ctccgcccc	ctgacgagca	tcacaaaaat	5340
cgacgctcaa	gtcagaggtg	gcgaaccgg	acaggactat	aaagatacca	ggcgtttccc	5400
cctggaagct	ccctcgtg	ctctcctgtt	cgcacccctg	cgcttaccgg	atacctgtcc	5460
gcctttctcc	cttcgggaag	cgtggcgctt	ctcatagct	cacgctgtag	gtatctcagt	5520
tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	aacccccctg	tcagcccgac	5580
cgctgcgcct	tatccggtaa	ctatcgtctt	gagtcacaac	cggttaagaca	cgacttatcg	5640

-80-

```

ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg cgggtgctaca 5700
gagttcttga agtggtggcc taactacggc tacactagaa ggacagtatt tgggtatctgc 5760
gctctctgta agccagttac cttcggaata agagttggta gctcttgatc cggcaaaaaa 5820
accaccgctg gtagcggtgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa 5880
ggatctcaag aagatccttt gatcttttct acgggggtctg acgctcagtg gaacgaaaaa 5940
tcacgttaag ggatttttgg catgagatta tcaaaaagga tcttcaccta gatcctttta 6000
aattaaaaat gaagttttaa atcaatctaa agtatatatg agtaaaactg gtctgacagt 6060
taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcattccata 6120
gttgccctgac tccccgctcg gtagataact acgatacggg agggcttacc atctggcccc 6180
agtgcctgcaa tgataccgcg agaccacgcg tcacgggctc cagatttatc agcaataaac 6240
cagccagccg gaagggccga ggcgagaagt ggtcctgcaa ctttatccgc ctccatccag 6300
tctattaatt gttgcccggg agctagagta agtagttcgc cagttaatag tttgcgcaac 6360
gttggttgcca ttgctacagg catcgtggtg tcacgctcgt cgtttgggtat ggcttcattc 6420
agctccgggt cccaacgac aaggcgagtt acatgatccc ccatgtttgtg caaaaaagcg 6480
gttagctcct tcggtcctcc gatcgtttgtc agaagtaagt tggccgcagt gttatcactc 6540
atgggttatgg cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct 6600
gtgactgggt agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc 6660
tcttgcccgg cgtcaataac ggataatacc ggcgccata gcagaacttt aaaagtgtctc 6720
atcattggaa aacgttcttc gggcgaaaaa ctctcaagga tcttaccgct gttgagatcc 6780
agttcgtatg aaccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc 6840
gtttctgggt aggcaaaaac aggaaggcaa atgtccgcaa aaaagggaat aagggcgaca 6900
cggaaatggt gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggg 6960
tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt 7020
ccgcgcacat ttccccgaaa agtgccacct gacgtagtta acaaaaaaaa gcccgccgaa 7080
gcgggcttta ttaccaagcg aagcgccatt cgccattcag gctgcgcaac tgttgggag 7140
ggcgatcggt gcgggcctct tocgctattac gccagctggc gaaaggggga tgtgctgcaa 7200
ggcgattaag ttgggtaacg ccagggtttt cccagtcacg acgttgttaa acgacggcca 7260
gtccgtaata cgactcactt aaggccttga cttagagggc gacggtatac agacatgata 7320
agatacattg atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt 7380
tgtgaaatth gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt 7440
ggggtggggc aagaactcca gcctgagatc cccgcgctgg aggatcatcc agccggcgctc 7500
cggaaaaacg attccgaagc ccaacctttc atagaaggcg gcggtggaat cgaaatctcg 7560
tagcacgtgt cagtctctgt cctcgccac gaagtgcacg 7600

```

<210> 116

<211> 7631

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo5'6XHS4eGFP Plasmid

<400> 116

```

cagttgcccg cggggtcgcg cagggcgaaac tccccccccc acggctgctc gccgatctcg 60
gtcatggccg gcccgaggc gtcccggaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacc caggccaggg tgttgtccg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagcccg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgccgcg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatata gaattcctgc 420
agcccccgcg atccgctcac caggacagcc ccccccaaaa gccccaggg atgtaattac 480
gtccctcccc cgtaggggg cagcagcgag ccgcccgggg ctccgctccg gtccggcgct 540
ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 600
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
acggggcgcg ggtccgctc acggggacag cccccccca aagccccag ggtatgtaatt 720
acgtccctcc cccgctaggg ggcagcagcg agccgcccgg ggctccgctc cggtcggcg 780
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc 840
acgggatcgc tttcctctga acgcttctcg ctgctctttg agcctgcaga cacctggggg 900
atacgggggc gcggtccgc tcacggggac agccccccc caaagcccc agggatgtaa 960
ttacgtccct ccccgctag ggggcagcag ctaggggaca ggggtcccg tccgggtccg 1020
cgctccccc gcattccccga gccggcagcg tgcggggaca gcccgggcac ggggaaggtg 1080
gcacgggata gctttcctc gaacgcttgc gctcagggg tgagcctgca gacacctggg 1140
ggatacgggg ctcccccgt ctcccccgt agggggcagc ccgagccccc ccagggatgt 1200
aattacgtcc ctcccccgt ccccccgt gagccggcag cgtgcccccc gctccgggtc 1260
ggcgtcccc ccgcatcccc ctgaacgctt cgtcgtgctc tttgagcctg cagcccgggc 1320
tggcacggga tcgcttctc cgcgtcacgg ggacagcccc ccccaaaagc cccaggggat 1440

```

-81-

gtaattacgt	ccctcccccg	ctagggggca	gcagcgagcc	gcccggggct	ccgctccggg	1500
ccggcgctcc	ccccgcaccc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
gggtggcagc	gatcgcttc	ctctgaacgc	ttctcgctgc	tctttgagcc	tgacagacac	1620
tgggggatac	ggggcccgcc	atccgctcac	ggggacagcc	ccccccaaa	gccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcat	ccccgagccg	gcagcggtgc	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggccccgctg	1980
gctgaccgcc	caacgacccc	cgccatttga	cgtaataaat	gacgtatgtt	cccatagtaa	2040
cgcaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggccccg	ctggcattat	gcccagtaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccc	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atttatattt	tttttaatta	2340
ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggggcggg	2400
cgagggcgcg	ggcgggcgga	ggcgggagag	ggcgggcgga	gccaatcaga	gcggcgcgct	2460
cgaaagtgtt	ccttttatgg	cgagggcgcg	ggcgggcgcg	ccctataaaa	agcgaagcgc	2520
gcgggcgggc	ggagtcgctg	cgttgccttc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccggcc	cggtctgac	tgaccgcgtt	actccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggcttgaatt	agcgcttggg	ttaatgacgg	ctcgtttctt	ttctgtgggt	2700
gcgtgaaagc	cttaaagggc	tccgggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgctgctgc	gtgtgtgtgc	gtggggagcg	ccgctgtgcg	ccgctgctgc	ccggcggtgc	2820
tgagcgctgc	ggggcgggcg	gggggctttg	tgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgccccggg	gctgcgaggg	gaacaaaggc	tgctgctggg	2940
gtgtgtgctg	gggggggtga	gcaggggggtg	tgggcgcggc	ggctcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggctt	cggtgctggg	gctccgtgcg	3060
gggctgtggc	cggggctcgc	cggtgccccg	ggggggtggc	ggcaggtggg	gggtgccccg	3120
ggggcggggc	cgctcgggc	cggggagggc	tcgggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcggtcg	ctgaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgctgag	3240
ggggcgaggg	acttctttg	tcccaaatct	ggcgagccg	aaatctggga	ggcgccgccc	3300
caccccctct	agcgggctgc	ggcgaaagcg	tgcgggcgcc	gcaggaagga	aatggcgggg	3360
gagggccttc	gtgctgctgc	gcgcccgcgt	ccccctctcc	atctccagcc	tcgggggtgc	3420
cgcaggggga	cggtgctctt	cgggggggag	ggggcagggc	gggggttcgg	ttctggcggt	3480
tgaccggcg	ctctagagcc	tcgtgtaacc	atgttcatgc	cttcttcttt	ttcctacagc	3540
tcctgggcaa	cgtgctggtt	gtgtgtgctg	ctcatcatct	tggcaaagaa	ttcgccacca	3600
tggtgagcaa	gggagaggag	ctgttcaccc	gggtggtgcc	catcctggtc	gagctggagc	3660
gcgacgtaaa	cggccacaag	ttcagcggtg	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagtct	atctgcacca	ccggcaagct	gcccgtgccc	tgggccaccc	3780
tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagtcc	gccatgcccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcgc	ccgaggtgaa	gttcgagggc	gacaccctgg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcataaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcggt	cagctcgccc	4140
accactacca	ccagaacacc	cccatcgggc	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	accccaacga	gaagcgcgat	cacatgggtc	4260
tgctggagtt	cgtgaccgcc	gcccggatca	ctctcgccat	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaagggtgt	ggctgggtgtg	gccaatgccc	4380
tggctcacia	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccttga	gcattctgact	tctggctaat	aaaggaaatt	tattttcatt	gcaatagtgt	4500
gttgaatttt	tttgtgtctc	tcactcgga	ggacatatgg	gagggcaaat	catttataaac	4560
atcagaatga	gtatttggtt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaagggtgg	tataaagagg	tcattcagtat	atgaaacagc	cccctgctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttctttaa	catccctaaa	catgttttac	cttctcttat	tagccagatt	tttctcctc	4800
tcctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	catgcctgca	ggtcgactct	agaggatccc	cgggtaccga	gctcgaattc	4920
gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	4980
catacgagcc	ggaagcataa	agtgtaaagc	ctgggggtgcc	taatgagtga	gctaactcac	5040
attaattgct	ttgcgctcac	tgcccgcctt	ccagtcggga	aacctgtcgt	gccagctgca	5100
ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgctg	attgggcgct	cttccgcttc	5160
ctcgctcact	gactcgctgc	gctcggtcgt	tcggctgcgg	cgagcggtat	cagctcactc	5220
aaaggcggtg	atacggttat	ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	5280
aaaaggccag	ccaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	5340
gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	5400
gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	tcctcctgct	gctcctcgtg	5460

-82-

```

tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct 5520
ttctcatagc tcacgcgtga ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg 5580
ctgtgtgcac gaaccccccg ttcagcccgga ccgctgcgcc ttatccggta actatcgctc 5640
tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat 5700
tagcagagcg aggtatgtag gcggtgctac agagtctctg aagtgggtggc ctaactacgg 5760
ctacactaga aggacagtat ttgggtatctg cgctctgctg aagccagtta ccttcggaaa 5820
aagagttggg agctcttgat ccggcaaaaca aaccaccgct ggtagcgggt gtttttttgt 5880
ttgcaagcag cagattacgc gcagaaaaaaa aggatctcaa gaagatcctt tgatcttttc 5940
tacggggtct atcaaaaagg ggaacgaaaaa ctcacgttaa gggatttttg tcatgagatt 6000
atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta 6060
aagtatatat gagtaaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat 6120
ctcagcgatc tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac 6180
tacgatacgg gagggttac catctggccc cagtgtgca atgataccgc gagaccacg 6240
ctcaccggct ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag 6300
tggtcctgca actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt 6360
aagtgttccg ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt 6420
gtcagctcgc tgcgtttggt tggcttcatt cagctccggg tcccaacgat caaggcgagt 6480
tacatgatcc cccatgttgt gcaaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt 6540
cagaagtaag ttggccgcag gatgcttttc catggttatg gcagcactgc ataattctct 6600
tactgtcatg ccatccgtaa gaccgagttg ctcttgcccg gactactcaa ccaagtcatt 6660
ctgagaatag tgtatgcggc taagaagtgt catcattgga aaacgttctt gggataatac 6720
cgcgccacat agcagaactt taaaagtgtc cagtccgatg taaccactc cggggcgaaa 6780
actctcaagg atcttaccgc ctctcaccag cgtttctggg tgagcaaaaa caggaaggca 6840
ctgatcttca gcatctttta taagggcgac acggaaatgt tgaatactca tactcttctc 6900
aaatgccgca aaaaagggaa tttatcaggg tttatcaggg atgagcggat atgagcggat 7020
ttttcaatat tattgaagca aatatggggt ttcggcgaca tttccccgaa aagtgccacc 7080
atgtatttag aaaaataaac agcccccgga agcgggcttt attaccaagc gaagcgccat 7140
tgacgtagtt aacaaaaaaa ctggtgggaa gggcgatcgg tgccggcctc ttccgctatta 7200
tcgccattca ggctgcgcaa atgtgtctga aggcgattaa gttgggtaac gccagggttt 7260
cgccagctgg cgaaaggggg aacgacggcc agtcctaat acgactcact taaggccttg 7320
tcccagtcac gacgttgtaa cagacatgat gatgagtttg gacaaaccac 7380
actagagggg cagtgaaaaa aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt 7440
aactagaatg ataagctgca ataaacaagt tggggtgggc gaagaactcc agcatgagat 7500
tgtaaccatt ccccgcgctg gaggcggcgt cccggaaaac gattccgaag cccaaccttt 7560
catagaaggc ggcgggtggaa tcgaaatctc gtagcacgtg tcagtcctgc tcctcggcca 7620
cgaagtgcac g 7631

```

<210> 117

<211> 4615

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo6XHS4 Plasmid

<400> 117

```

cagttgccgg ccgggtcgcg caggggcgaac tcccgcccc acggctgctc gccgatctcg 60
gtcatggccg gcccggaggc gtcccgggaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacacc caggccaggg tgttgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg togtcccgga ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagccgg tcgggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgcatgc ctgcaggctg 420
actctagtgg atcccccgcc cgtatcccc caggtgtctg caggctcaaa gagcagcgag 480
aagcgttcag aggaaagcga tcccggtgcca ccttccccgt gcccgggctg tccccgcacg 540
ctgccggctc ggggatgcgg ggggagcgcc ggaccggagc ggagccccgg gcggctcgct 600
gctgccccct agcggggggag ggacgtaatt acatccctgg gggctttggg gggggcgctg 660
ccccgtgagc ggatcccgcg ccctgtatcc cccaggtgtc tgagggtca aagagcagcg 720
agaagcgttc agaggaaagc gatcccgctg cacttcccc gtgcccgggc tgtccccgca 780
cgctgccggc tcggggatgc ggggggagcg cccgaccgga gcggagcccc gggcggtctg 840
ctgctgcccc ctagcggggg agggacgtaa ttacatccct gggggctttg ggggggggct 900
gtccccgtga cggcggcgcc gggcccgctat ccccaggtg tctgcaggct caaagagcag 960
cgagaagcgt tcagaggaaa gcatcccggt gccaccttcc ccgtgcccgg gctgtccccg 1020
cacgctgccg gctcggggat gcggggggag cgcgggaccc gagcggagcc cgggggggct 1080
cgctgctgcc ccctagcggg ggaggagcgt aattacatcc ctgggggctt tggggggggg 1140
ctgtccccgt gagcggatcc gcggccccgt atcccccagg tgtctgcagg ctcaaagagc 1200

```


-83-

```
agcgagaagc gttcagagga aagcgatccc gtgccacctt ccccgtagcc gggctgtccc 1260
cgcacgctgc cggctcgggg atgccccggg agcgccggag cggagcggag ccccgggcgg 1320
ctcgtctgtg cccctagcgg ggggaggggac gtaattacat ccctggggggc tttggggggg 1380
ggctgtcccc gtgagcggat ccgccccccc gtatccccca ggtgtctgca ggctcaaaga 1440
gcagcgagaa gcgttcagag gaaagcgatc ccgtgccacc ttccccgtgc ccgggctgtc 1500
cccgacgctg gccggctcgg ggatgccccg ggagcgccgg accggagcgg agccccgggc 1560
ggctcgtctg tgccccctag cgggggaggg acgtaattac atccctgggg gctttggggg 1620
ggggctgtcc ccgtgagcgg atcccgggcc ccgtatcccc caggtgtctg caggtcaaaa 1680
gagcagcgag aagcgttcag aggaaagcga tccccgtcca ccttccccgt gcccgggctg 1740
tccccgcacg ctgccggctc ggggatgcgg ggggagcgcc ggaccggagc ggagccccgg 1800
gccccgtcgt gctgcccctc agcgggggag ggacgtaatt acatccctgg gggctttggg 1860
ggggggctgt ccccgtagc ggatcccgcg ggctgcagga attcgtaatc atgggtcatag 1920
ctgtttcctg tgtgaaattg ttatccgctc tgcctaatac acaattccac acaacatacg agccggaagc 1980
ataaagtgtg aagcctgggg tgccctaatg gtgagctaac tcacattaat tgcgttgctc 2040
tactgccccg ctttccagtc gggaaacctg tctgcccagc tgcattaatg aatcgccaa 2100
cgcgcgggga gaggcggttt gcgtattggg cgctcttccg cttcctcgtc cactgactcg 2160
ctgctcagc ctgctcggct cgccgagcgc gtatcagctc actcaaaggc ggtaatacgg 2220
ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag 2280
gccaggaacc gtaaaaaggc cgctgtgctg gcgtttttcc ataggctccg cccccctgac 2340
gagcatcaca aaaatcgacg ctcaagtcag agtgggcgaa acccgacagg actataaaga 2400
taccagcggt ttccccctgg aagctccctc gtcgctctc ctgttccgac cctgcccgtt 2460
accggatacc tgtccgctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc 2520
tgtaggtatc tcagttcggg gtaggtcgtt cgctccaagc tgggctgtgt gcacgaacc 2580
ccccgtcagc ccgaccgtgc cgcttatcc ggtaaactatc gtcttgagtc caaccggta 2640
agacacgact tatcgccact ggcagcagcc actggttaaca ggattagcag agcgaggat 2700
gtaggcgggt ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca 2760
gtatttggtg tctgctctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct 2820
tgatccggca aacaaaccac cgctggtagc ggtgggtttt ttggttgcaa gcagcagatt 2880
acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct 2940
cagtggaacg aaaaactcag ttaagggatt ttggtcatga gattatcaaa aaggatcttc 3000
acctgatccc ttttaatta aaaaatgaagt tttaaatcaa tctaaagtat atatgagtaa 3060
acttggtctg acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta 3120
tttcgttcat ccatagttgc ctgactcccc taactacgat acggggagggc 3180
ttaccatctg gccccagtgc tgcaatgata cgcgcagacc cagctcacc ggctccagat 3240
ttatcagcaa taaaccagcc agccggaagg cccgagcgca gaagtgggtc tgcaaaatga 3300
tccgcctcca tccagctctat taattgttgc cgggaagcta gagtaagtag ttcgccagtt 3360
aatagtttgc gcaacgttgt tgccattgct tgggtgtcacg ctcgctcgtt 3420
ggtagtgctt cattcagctc cggttcccaa gagttacatg atcccccatg 3480
ttgtgcaaaa aagcggttag ctccctcggt ttgtcagaag taagttggcc 3540
gcagtggtat cactcatggt tatggcagca ctgtcataatt ctcttactgt catgccatcc 3600
gtaagatgct tttctgtgac tggtgagtag tcaaccaagt cattctgaga atagtgtatg 3660
cggcgaccga ttctctcttg cccggcgta ataccgggata ataccgcgcc acatagcaga 3720
actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta 3780
ccgctgttga gatccagttc gatgtaacct ccaactgatc ttcagcatct 3840
tttactttca ccagegtttc tgggtgagca aaaaacaggaa ggcaaaatgc cgcaaaaaag 3900
ggaataaggg cgacacggaa atggtgaata ctcatactct tcctttttca atattattga 3960
agcatttatc agggttattg totcatgagc ggatacatat ttgaatgtat ttagaaaaat 4020
aaaaagcccc ccgaagcggg cacttaagtc cgaaaagtgc cacctgacgt agttaacaaa 4080
gcaactgttg ggaagggcga tccgtgcggg aagcgaagcg ccattcgcca ttcaggctgc 4140
ggggatgtgc tgcaaggcga ttaagttggg cctcttcgct attacgccag ctggcgaaag 4200
gtaaaaacgac ggccagtcg taatacgact taacgccagg gttttccag tccagcagtt 4260
tatacagaca tgataagata cattgatgag cacttaaggc cttgactaga gggtcgacgg 4320
aaaaaatgct ttatttgtga tttggacaaa ccacaactag aatgcagtga 4380
tgcaataaac aagttggggg gggcgaagaa gctattgctt tatttgtaac cattataagc 4440
catccagcgg cgtccccgga cccagcatg agatccccgc gctggaggat gctggaggat 4500
ggaatcgaaa tctcgtagca cgtgtcagtc ctccagcaac ctttcataga agggcgcggt 4560
gccacgaagt gcacg 4615
```

<210> 118
<211> 17384
<212> DNA
<213> Artificial Sequence

<220>
<223> pFK161 Plasmid

<400> 118

-84-

gcgcacgagg	gagcttccag	ggggaaacgc	ctgggtatctt	tatagtcctg	tccggggttc	60
gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	aggggggcgg	agcctatgga	120
aaaacgccag	caacgcggcc	tttttacggg	tcctggcctt	ttgctggcct	tttgctcaca	180
tggtttttcc	tgcggtatcc	cctgattctg	tggataaccg	tattaccgcc	tttgagttag	240
ctgataccgc	tcgcgcgcag	cgaacgaccg	agcgcgcgca	gtcagtgagc	gaggaagcgg	300
aagagcgctg	acttccgcgt	ttccagactt	tacgaaacac	ggaaaccgaa	gaccattcat	360
gttggttgctc	aggtgcgcga	cgttttgtag	cagcagtcgc	ttcacgttcg	ctcgcgtatc	420
ggtgattcat	tctgctaacc	agtaaggcaa	ccccgccagc	ctagccgggt	cctcaacgac	480
aggagcacga	tcatgcgcac	cgcgcagatc	cagacatgat	aagatacatt	gatgagtttg	540
gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	600
ttgctttatt	tgtaaccatt	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	660
atthttatgtt	tcagggttcag	ggggagggtg	ggggagggtt	ttaaagcaag	taaaacctct	720
acaaatgtgg	ttaggtgat	aaaagctaaa	ggtacacaat	ctatacatca	aatattcctt	780
attaaccctc	ctatgcctgt	gtggagtaag	aaaaaacagt	ttttgagcat	agttattaat	840
agcagacact	ataaagggtta	cagaatattt	ttccataatt	atgttatgat	tataactggt	900
ttgctacttt	tgaggtgtaa	tagcaaaagca	agcaagagtt	ttcttgata	gcagtgagc	960
actcaaaaaa	cttagcaatt	ctgaaggaaa	gtccttgggg	ctattactaa	acacagcatg	1020
ttttgggagga	gtagaatgtt	gagagtcagc	agtagcctca	tcttctacct	ttctcttctt	1080
ttctgagcaa	aacagggttt	cctcattaaa	ggcattccac	tcatcactag	atggcatttc	1140
tccataggtt	ggaatctaaa	atacacaaac	aattagaatc	cactgctccc	attcatcagt	1200
acttaaaaaa	tttatattta	ccttagagct	ttaaatctct	agtagtttaa	cacattatac	1260
tgtcacacca	cagaagtaag	gttccctcac	aaagatccgg	gtaggtagtt	tgtccaatta	1320
tccccactcc	tgaggttcgg	gggcatggat	gcgcggatag	accaaagcgg	ccatcgtgcc	1380
gccgacggat	ttgactgcc	ggtagaactc	cgcggatag	ccgctgctgg	tttccctggat	1440
aaccaactcg	cagggggatc	gagcccgggg	gcgaggtcgt	ccagcctcag	gcagcagctg	1500
cgcgtggagga	tcataccagc	ggcgctccgg	ttgggcgaaga	actccagcat	gagatccccg	1560
aaggcgccgg	tggaatcgaa	atctcgtgat	aaaacgattc	cgaagcccaa	ctttctatag	1620
tcgaacccca	gagtcccgt	cagaagaact	ggcaggttgg	gcgtcgcttg	gtcgggtcatt	1680
gcgaatccgg	agcggcgata	ccgtaaaagca	cgtcaagaag	gcgatagaag	gcgatgcgct	1740
gctcttcagc	gtagccaacg	ctatgtcctg	cgaggaagcg	gtcagcccat	tcgcccgcga	1800
gccggccaca	gtcgatgaat	ccagaaaagc	ctatgtcctg	atagcgggtc	gccacaccca	1860
aggcatcgcc	atgggtcacg	acgagatcct	ggccattttc	caccatgata	ttcggcaagc	1920
gaacagttcg	gctggcgcgga	gcccctgatg	cgccgtcggg	atgcgcgcct	tgagcctggc	1980
accggcttcc	atccgagtag	gtgctcgttc	cttctcgtcc	agatcatcct	gatcgacaag	2040
gcaggtagcc	ggatcaagcg	tatgcagccg	gatgcgatgt	ttcgcttggt	ggtcgaatgg	2100
ctcggcagga	gcaaggtgag	atgacaggag	ccgcattgca	tcagccatga	tggtactttt	2160
ccagtcctct	cccgcttcag	tgacaacgtc	atcctgcccc	ggcacttcgc	ccaatagcag	2220
ggccagccac	gatagcccg	ctgcctcgtc	gagcacagct	gcgcaaggaa	cgcccgctgt	2280
gggtcttgaca	aaaagaaccg	ggcgcccctg	ctgcagttca	ttcagggcac	cggacagggtc	2340
gcagccgatt	ttgtgtgtg	cccagtcata	cgttgacagc	cggaaacacg	cggcatcaga	2400
agaacctgcg	tgcaatccat	cctgttcaat	cccgaatagc	ctctccacc	aagcggccgg	2460
atcagatctt	gatcccctgc	gccatcagat	cattgcgaaac	gatcctcatc	ctgtctcttg	2520
tttgccaggg	ttcccaacct	taccagaggg	ccttggcggc	aagaaagcca	tcagttttac	2580
tgtccataaa	accgccagat	ctagctatcg	cgccccagct	ggcaattccg	gttcgcttgc	2640
tctctttgcg	cctgcgtttt	cccttgtcca	ccatgtaagc	ccactgcaag	ctacctgctt	2700
tcagcaccgt	ttctgcggac	tggttttcta	gatagcccag	tagctgacat	tcatccgggg	2760
ccctgagtg	ttgcggcagc	gtgaaaagct	cgtgttcgcg	ttccttttag	agcccttgcg	2820
tcctcactac	ttctggaata	gctcagaggc	cttgcaaaaag	cctaggcctc	caaaaaagcc	2880
gccatggggc	ggagaatggg	cggaaactggg	cgaggcgggc	taaaataaaaa	aaattagtca	2940
gggcgggact	atggttgctg	actaattgag	cggagttagg	ggcgggatgg	gcggagttag	3000
ggagcctggg	gactttccac	acctgggtgc	atgcatgctt	tgcatacttc	tgccgtgctg	3060
tctgcctgct	ggggagcctg	gggactttcc	tgactaattg	agatgcatgc	tttgcatact	3120
gatctgcagg	acccaacgct	gcccagatg	acaccctaac	tgacacacat	tcacacagccg	3180
gcgatggata	tggtctgcca	agggttggtt	cgcgcgctgc	ggctgctgga	gatggcggac	3240
ttggctccaa	ttcttgaggt	ggtgaatccg	tgcgcaattca	cagttctccg	caagaattga	3300
tcgaggtggc	ccggctccat	gcaccgcgac	ttagcgaggt	gccgccggct	tcatttcagg	3360
cggcgccctac	aatccatgcc	aaccggttcc	gcaacgcggg	gaggcagaca	aggtataggg	3420
gacgatcagc	ggtccaatga	tcgaagttag	atgtgctcgc	cgaggcgcat	aatcgccgt	3480
ctgtccctga	tggtcgatcat	ctacctgcct	gctggtaaga	gccgcgagcg	atccttgaag	3540
atgcccgcgg	aagcgagaag	aatcataatg	ggacagcatg	gcctgcaacg	cggcatcccg	3600
gccagcaaga	cgtagcccag	cgcgtcgggc	gggaaggcca	tcagccctcg	cgtcgcgaa	3660
gccgaaacgt	ttggtggcgg	gaccagtgac	gcgaataatg	gcgataatgg	cctgcttctc	3720
gaataccgca	agcgacaggg	cgatcatcgt	gaaggcttga	gcgagggcgt	gcaagattcc	3780
aatgacccag	agcgctgccc	gcacctgtcc	cgcgctccag	cgaagcggtt	cctcgccgaa	3840
aagtgcggcg	acgatagtc	tgccccgcg	tcagagttgc	atgataaaga	agacagtc	3900
tctcaagggc	atcggtcgac	gctctccctt	ccaccggaag	gagctgactg	ggttgaaggc	3960
			atgcgactcc	tgcattagga	agcagccag	4020

-85-

tagtaggttg	agggcgttga	gcaccgcccgc	cgcaagggaat	ggtagcatgca	aggagatggc	4080
gcccacacagt	cccccgggcca	cgggcctgccc	accatacccca	cgccgaaaca	agcgctcatg	4140
agcccgaaagt	ggcgcagcccgc	atcttccccca	tccggtgatgt	cggcgatata	ggcgccagca	4200
accgcacctg	tggcgccgggt	gatgcccggcc	acgatgcgtc	cggcgtagag	gatcttggca	4260
gtcacagcat	gcgcataatcc	atgcttcgcac	catgcgctca	caaagttaggt	gaatgcgcaa	4320
tgtagtacc	acatcgctcat	cgctttccac	tgctctcgcg	aataaagatg	gaaaatcaat	4380
ctcatggtaa	tagtccatga	aaatccttgt	attcataaat	cctccaggta	gctatatgca	4440
aattgaaaca	aaagagatgg	tgatctttct	aagagatgat	ggaatctccc	ttcagtatcc	4500
cgatgggtcaa	tgcgtggat	atgggataga	tgggaatatg	ctgattttta	tgggacagag	4560
ttgcgaactg	ttcccaacta	aaatcatttt	gcacgatcag	cgcactacga	actttaccca	4620
caaatagtca	ggtaataaat	cctgatataa	agacagggtg	ataaatcagt	cttctacgcg	4680
catcgacgcg	gcacaccgta	gaaagtcttt	cagtttgtgag	cctggggcaaa	ccgttaactt	4740
tcggcggttt	tgcgtgcca	caggctcacg	tctaaaagga	aataaatcat	gggtcataaa	4800
attatcacgt	tgtccggcgc	ggcgacggat	gttctgtatg	cgctgttttt	ccgtggcgcg	4860
ttgctgtctg	gtgatctgcc	ttctaaatct	ggcacagccg	aattgcgcga	gcttgggttt	4920
gctgaaacca	gacacacagc	aactgaatac	cagaaagaaa	atcactttac	ctttctgaca	4980
tcagaagggc	agaaatttgc	cgttgaacac	ctggtcaata	cgctgttttg	tgagcgagca	5040
tattgcgctt	cgatgacgct	tggcggtgag	attgatacct	ctgctgcaca	aaaggcaatc	5100
gacgagctgg	accagcgcat	tgcgtgacacc	gtctccttcg	aacttattcg	caatggagtg	5160
tcattcatca	aggacgcccgc	tatcgcaaat	gggtgctatcc	acgcagcggc	aatcgaaaca	5220
cctcgccggg	ctacaacatc	agccttggta	tccagcgtga	tccagcgtga	tgagccagcg	5280
cagaacaagg	taaccgctcag	tgcggataag	ttcaaagtta	aacctggtgt	tgataccaac	5340
attgaaacgt	tgatcgaaaa	cgcgctgaaa	aacgctgctg	aatgtgcggc	gctggatgtc	5400
acaaagcaaa	tggcagcaga	caagaaaagcg	atcgatgaac	tggcttcccta	tgtccgcacg	5460
gccatcatga	tggaaatgttt	ccccgggtggt	gttatctggc	agcagtgcgg	tcgatagtat	5520
gcaattgata	attattatca	tttgccgggtc	ctttccggcg	atccgccttg	ttacggggcg	5580
gcgacctcgc	gggttttcgc	tatttatgaa	aattttccgg	tttaaggcgt	ttccgttctt	5640
cttcgtcata	acttaatgtt	tttattttaa	ataccctctg	aaaagaaagg	aaacgacagg	5700
tgctgaaagc	gagctttttg	gcctctgtcg	tttcccttct	ctgtttttgt	ccgtggaatg	5760
aacaatggaa	gtcaacaaaa	agcagctggc	tgacattttc	ggtagcgagta	tcctgtaccat	5820
tcagaactgg	caggaaacagg	gaatgcccg	tctgcgaggg	ggtagcagg	gtaatgaggt	5880
gctttatgac	tctgcgcgcg	tcataaaatg	gtatgccgaa	agggatgctg	aaattgagaa	5940
cgaaaagctg	cgccgggagg	ttgaagaact	gcccgcaggcc	agcgaaggcag	atccacagga	6000
cggtgtgggt	cgccatgatc	gcgtagtcca	tagtggctcc	aagtagcgaa	gcgagcagga	6060
ctgggcccgcg	cggagcgggt	tcggagacgc	tccgagaacg	ggtagcgata	gaaattgcat	6120
caacgcataat	agcgtagca	gcacgccata	gtgactggcg	atgctgtcgg	aatggacgat	6180
atcccgcaag	agggccggca	gtaccggcat	aaaccaagcct	atgcctacag	catccagggt	6240
gacgggtgccg	aggatgacga	tgagcgcatt	gttagatttc	atacacgggtg	cctgactgcg	6300
ttagcaatatt	aactgtgata	aactaccgca	ttaaagctta	tcgatgataa	gcgggtcaaac	6360
atgagaattc	gcggccgctc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccatc	6420
cgtctgccgg	tgggtgtgtg	aaggcagggg	tgcggctctc	cggcccgacg	ctgcccgcgc	6480
cgcacttttc	tcagtgggtc	gcgtgggtcc	tgctggatgtg	tgaggcgccc	gggtgtgccc	6540
tcacgtgttt	cactttgggtc	gtgtctcgct	tgaccatgtt	cccagagtcg	gtggatgtgg	6600
ccggtggcgt	tgcataccct	tcccgctctg	tggtgtgcacg	cgctgtttct	tgtaagcgtc	6660
gaggtgtctcc	tggagcgttc	caggttttgtc	tcttaggtgc	ctgcttctga	gtgggtgggt	6720
gcgctcccca	ttccctgggtg	tgcctccgggt	gctccgtctg	gctgtgtgcc	ttcccggtttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgcccac	cccgcgacta	gtacgcctgt	gcgtagggtc	gggtgctgagc	6900
ggtcgcggct	ggggttggaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccgggt	gcagggtctc	ccctgtccgc	ggatgtctag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcgg	cgcccgcggt	7080
ttggtcttct	ggtttccctg	tgtgctcgtc	gcatgcatcc	tctctcgggtg	gcccgggctc	7140
gtcgggggtt	tgggtccgctc	ccgcctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaagggtg	cgggtctctt	acggtctcga	ggggtctctc	ccgaatgggtc	ccctggaggg	7260
ctcgccccct	gaccgcctcc	cgcgcgcgca	gcgtttgctc	tctcgtctac	cgcggccgcg	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtggttgg	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttcctcc	cctcctgagg	gcccgcgtgc	ggacgggggtg	7500
tgggtaggcg	acgggtgggt	cccgggtccc	caccgctctt	ccggtgcctc	aoccgctcct	7560
tcggtcgcgt	gcgtccctct	cgctcgcgtc	cacgactttg	gcgctcccgc	cgacggcgggc	7620
ctgcgcgcgc	cgtggtgcgt	gctgtgtgct	tctcgggctg	tgtggttgtg	tcgcctcgcc	7680
cccccttccc	cgcggcagcg	ttcccacggc	tggcgaaate	gcgggagtc	tccttcccct	7740
cctcgggggtc	gagaagggtc	gtgtctggcg	ttgatgtatc	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaacggctg	ttggccgcgt	ccggcgcgac	gtcggacgtg	gggacccact	7860
gcgctcggg	ggtcttcgtc	ggtaggcatc	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cgggtgcgcc	tcctcgggct	cccggggggc	cgtcgtgttt	cgggtcgggt	cggcgctgca	7980
ggtgtggtg	gactgctcag	gggagtggtg	cagtgtgatt	cccgcgggtt	ttgcctcgcg	8040

-86-

tgccctgacc	ggtccgacgc	ccgagcggtc	tctcgggtccc	ttgtgaggac	ccccttccgg	8100
gagggggccc	tttcggccgc	ccttgccgcg	gtcgcgggcc	ctcgttctgc	tgtgtcggtc	8160
ccccctcccc	gctcgcgcga	gcgggtcttt	tttcctctct	ccccccctct	cctctgactg	8220
accctgtggc	gtgctgtcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgtccggg	8280
cggtcaccgg	ggtcttgggg	gggggcccag	gggtaaagaaa	gtcggctcgg	cgggcgggag	8340
gagctgtggt	ttggagggcg	tccccggccc	gcggccgtgg	cggtgtcttg	cgcggctctg	8400
gagagggctg	cgtgcgaggg	gaaaagggtg	ccccgcgagg	gcaaagggaa	agaggctagc	8460
agtggtcatt	gtcccgacgg	tgtgggtggc	tgttggccga	ggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgcg	gttggggcct	gcgggagtgc	cgagggtggg	8580
accctggcgg	ttgggattaac	cccgcgcgcg	tgtcccggtg	tggcggtggg	ggctccggtc	8640
gatgtctacc	tccctctccc	cgaggtctca	ggccttctcc	gcgcgggctc	tcggccctcc	8700
cctcgttctc	ccctctcgcg	gggttcaagt	cgctcgtcga	cctccccctc	tcctgtcttc	8760
catctctcgc	gcaatggcgc	cgcccgagtt	cacggtgggt	tcgtcctccg	cctccgcttc	8820
tcggccgggg	ctggccgctg	tccggtctct	ccgcgccgac	cccgttggc	gtggctcttc	8880
ctcgcgggct	tcgcggactc	ctggcttcgc	ccggagggtc	agggggcttc	ccggttcccc	8940
gacgttgcgc	ctcgtgtctg	tgtgcttggg	gggggcccgc	tgccgctcct	gcccgcgggt	9000
gagccctcgc	cgcaccgcgc	ggtgtgcggt	ttcgcgcgcg	ggtcagttgg	gcccgtggcg	9060
tgtgtcgcgt	cgggagcgtg	tccgcctcgc	ggcggtctaga	cgcggtgtgc	gccgggctcc	9120
gacgggtggc	ctatccaggg	ctcgcgcccg	ccgacccccg	cctgcccgtc	ccggtgggtg	9180
tcgttggtgt	ggggagtgaa	tggtgtctac	ggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcccgtgt	cgcgcttctc	ttccgcgcaa	ccccacgccc	aaccaccac	cctgctctcc	9300
cggcccgggtg	cggtcgacgt	tccggctctc	ccgatgccga	gggggttcggg	atttgtgccc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
gttggtcttg	cccgctgcgt	gtgctcgcgg	acgggttttg	tcggaccccc	acggggctcg	9480
tccggccgca	tgcactctcc	cgctccgcgc	gagcgcggcg	ccggctcacc	cccggtttgt	9540
cctccccgga	ggctctccgc	cgccgcgcgc	tcctcctcct	ctctcgcgct	ctctgtcccg	9600
cctggtcctg	ttccaccccc	gacgctccgc	tcgcgcttcc	ttacctggtt	gacccctcca	9660
ggtagcatat	tcttctctca	aagattaaag	atagcatgtc	taagtacgca	cgcccggtac	9720
agtgaactcg	cgaatggctc	attaaatcag	ttatggttcc	tttggtcgct	cgctcctctc	9780
ctacttggtg	aactgtggta	attctagagc	taatacatgc	cgacggggcg	tgacccccct	9840
tcccgggggg	ggatgcgtgc	atttatcaga	cccaaaccac	cccggtgagc	tccctccggg	9900
ctccggccgg	gggtcgggcg	ccggcggtct	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccg	tggcgggcga	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgc	taccatgggtg	accacgggtg	acggggaatc	aggggttcgat	tcggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactccc	10140
cgacccgggg	aggtagtgac	gaaaaataac	aatacaggac	tctttcgagg	ccctgtaatt	10200
ggaatgagtc	cacttttaaat	cctttaacga	ggatccattg	gagggcaagt	ctggtgccag	10260
cagccggcgt	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcgggcggtc	cgccgcgagg	cgagtaccgc	cccgtccccg	10380
ccccttgccct	ctcggcgccc	cctcgatgct	cctagctgag	tgtcccgcgg	ggcccggaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccga	gcccgcctgga	taccgcagct	10500
aggaataatg	gaataggacc	gcggttctat	tttgttggtt	ttcggaactg	aggccatgat	10560
taagaggggac	ggccggggggc	attcgtattg	cgccgctaga	ggtgaaattc	ttggaccggc	10620
gcaagacgga	ccagagcgaa	agcatttgcc	aagaattggtt	tcattaatca	agaacgaaag	10680
tcggaggttc	caagacgata	agataccgct	gtagttccga	ccataaacga	tgccgactgg	10740
cgatcgggcg	gcgttatttc	catgacccgc	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggt	tgcaaaagctg	aaacttaag	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccgga	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggctg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgcacgcgcg	ctacactgac	tggctcagcg	tgtgcctacc	ctgcgcggcg	aggcgcggtg	11220
aacccgttga	accccatctg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtcctgccc	ctttgtacac	11340
accgcccgtc	gctactaccg	attggatggt	ttagtggagc	cctcggatcg	gccccgcccg	11400
ggtcggccca	cggccctggc	ggagcgctga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagtc	gtaacaaggt	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcgggcg	ggccgcgtct	ccccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgcgtgcgtc	ccgggtcccg	tcgcccgcgt	gtggagcgag	gtgtctggag	tgaggtgaga	11640
gaagggggtg	gtggggctcg	tctgggtccg	tctgggacgg	cctccgattt	cccccccccc	11700
tccctctccc	ctcgtccggc	ctgacctccg	ccaccctacc	gcggcgggcg	ctgctcggcg	11760
gcgtcttgcc	tctttcccgt	ccggtctctc	cgtgtctacg	aggggcggga	cgtcgttacg	11820
ggtttttgac	ccgtcccggg	ggcgttcggg	cgtcggggcg	cgcgctttgc	tctccgggca	11880
cccattcccc	ccgcggctct	ggcttttcta	cgttggttgg	ggcggttggtc	gcgtgtgggg	11940
ggatgtgagt	ctcgcgtgtg	ggctcgcggc	tcccgatgcc	acgcttttct	ggcctcgcgt	12000
gtcctccccg	ctcctgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaagga	12060

-87-

ccccgggggg	gtcgccctgc	cgccccagg	gtcggggggc	gggtggggccc	gtaggggaagt	12120
cgggtcgttcg	ggcggtcttc	cctcagactc	catgaccctc	ctccccccgc	tgccgcccgtt	12180
cccgaaggcgg	cggctcgtgtg	gggggggtgga	tgcttgaggc	ccctcgggc	gccgtggggg	12240
cccgaaccgc	ggcgccggct	tgcccgattt	ccgcgggtcg	gtcctgtcgg	tgccgggtcgt	12300
gggttcccgt	gtcgttcccg	tgtttttccg	ctcccgaccc	tttttttttc	ctcccccca	12360
cacgtgtctc	gtttcgttcc	tgctggccgg	cctgaggcta	ccctcggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcgggtgg	tcgttggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	ctcacacccg	aaataccgat	acgactctta	gcgggtggatc	actcggctcg	12540
tgctcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttctctc	cggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcacccgc	tgcggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	accggggtcg	ggccctccgt	ctcccgaaat	tcagacgtgt	12780
ggggcggttgt	cgggtgtggcg	cgcgcgcccg	cgtcgcggag	cctgggtctcc	cccgcgcac	12840
cgcgtcgcgc	gcttcttccc	gctcgcggtg	tcccgcctc	gcccgtgcac	cccggctcgt	12900
gcctcgcgtc	ggcgccctccc	ggaccgctgc	ctcaccagtc	tttctcgggtc	ccgtgccccg	12960
tgggaaacca	cgcgcgcccc	gtggcgcccc	gggggtggcg	cgtccgcac	tgctctgggtc	13020
gaggttggcg	cttgagggtg	tgctgcgcgc	gaggtgggtg	tcgggtccct	gcggccggcg	13080
gggttgcggg	gtggcggtcg	acgaggggcg	gtcggtcgcc	tgcggtgggt	gtctgtgtgt	13140
gtttgggtct	tgcgctgggg	gaggcggggg	cgaccgctcg	cgggggtggc	gcggctcgccc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatcccggg	tggcgttgcg	agggagggtt	tgccgtcccg	cgtccgtccg	tcctccctc	13320
cctcgggtggg	cgccttcgcg	ccgcacgcgg	ccgctagggg	cggctcggggc	ccgtggcccc	13380
cgtggctctt	ctcgtctctc	gcttctcctt	caccggggcg	gtaccgcgtc	cggcgccggc	13440
ccgcgggacg	cgcggcgctc	cgtgcgcgga	tcgcagtcac	cccgggtgt	tcgagttcg	13500
gggagggaga	gggcctcgct	gacccggtgc	gtcccggctt	ccctgggggg	gaccggcggt	13560
ctgtgggctg	tgctcccggg	gggttgctgt	tgagtaagat	cctccacccc	cgccgcccctc	13620
ccctcccggc	ggcctctcgg	ggacccctcg	agacgggtcg	ccggctcgtc	ctcccgtgcc	13680
gccgggtgctg	gtctctttcc	cgcgcgctc	ctcgtctctc	tttcccgcg	gctgggcgcg	13740
tgtccccct	ttctgaccgc	gacctcagat	cagacgtggc	gaccgcgtga	atttaagcat	13800
attagtacg	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgccgaatc	ccgcgcgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gaccactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gccagccccg	13980
tggaacggtg	gaggccggta	gcggccccgg	cgcgcggggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaacctca	tctaaggcta	aataccggca	14100
cgagaccgat	agtcacaag	taccgtaagg	gaaagtgtga	agaactttg	aagagagagt	14160
tcaagagggc	gtgaaaccgt	taagaggtaa	acgggtgggg	tcgcgcagat	ccgcccggag	14220
gattcaaccc	ggcggcgcgc	gtccggccgt	gcccggtggg	ccggcggtat	ctttcccgtc	14280
ccccgttctc	cccgaccctc	ccaccgcgc	gtcgttcccc	tttccctccc	cgcgtccggc	14340
gcctccggcg	gcgggcgcgg	gggtgggtgt	gggtgggtgg	cgcgggcggg	gccgggggtg	14400
gggtcgcgcg	gggaccgccc	ccggccggcg	accggccgcc	gccggggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tcgggacggg	ccgggaaggc	ccggtggggg	aggtggctcg	14520
gggggggcgg	gcgctctcag	ggcgcgccga	accactcac	cccgagtggt	acagccctcc	14580
ggccgcgctt	tcgccgaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgtctctc	14640
ctctcccccc	gtccgcctcc	cgggcggggc	tgggggtggg	ggccggggcg	ccctcccac	14700
ggcgcgaccg	ctctcccacc	ccctccgctc	gctcctctcg	gggcccgggtg	ggggggcggg	14760
cggactgtcc	ccagtgcgcc	ccgggcgtcg	tcgcgcctcg	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcagaa	gccgagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgaccgtctc	tgaaacacgg	accaaggagt	ctaacgcgtg	cgcgagtcag	gggctcgtcc	14940
gaaagccgcc	gtggcgcaat	gaagggtgaag	ggccccgccc	ggggggccga	ggtgggatcc	15000
cgaggcctct	ccagtccgcc	gaggcgccac	caccggcccc	tctcgcccg	cgcgcggggg	15060
agggtggagca	cgagcgtacg	cgttaggacc	cgaaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggagggtc	gtagcgggtc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaaggc	15300
aatgattaga	ggctctgggg	ccgaaacgat	ctcaacctat	tctcaaactt	taaatgggta	15360
agaagcccg	ctcgctggcg	tggaacgggg	cgtggaatgc	gagtgcctag	tgggccactt	15420
ttggttaagca	gaactggcgc	tgccggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgtcatc	agaccccgaa	aaaggtgttg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtcggaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccgcccgctc	gccgcagtcg	gaacggaaacg	15660
ggacgggagc	ggcgcgcaat	tcttgaagac	gaaagggcct	cgtgatacgc	ctatttttat	15720
aggttaatgt	catgataata	atggtttctt	agacgtcagg	tggcactttt	cgggggaaatg	15780
tgccggaac	ccctatttgt	ttatttttct	aaatacattc	aaatatgtat	ccgctcatga	15840
gacaataacc	ctgataaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	15900
atttccgtgt	cgccttattt	cccttttttg	cggcattttg	cttccgtgtt	ttgctcaacc	15960
agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	ggtgcacgag	tgggttacat	16020
cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	cgcgccgaag	aacgttttcc	16080

-88-

aatgatgagc	acttttaaag	ttctgctatg	tggcgcggtg	ttatcccgtg	ttgacgcggg	16140
gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	gacttggttg	agtactcacc	16200
agtcacagaa	aagcatctta	cggatggcat	gacagtaaga	gaattatgca	gtgctgccat	16260
aaccatgagt	gataacactg	cggccaactt	acttctgaca	acgatcggag	gaccgaagga	16320
gctaaccgct	tttttgcaca	acatggggga	tcatgtaact	cgccttgatc	gttgggaacc	16380
ggagctgaat	gaagccatac	caaacgacga	gcgtgacacc	acgatgcctg	cagcaatggc	16440
aacaacggtg	cgcaaaactat	taactggcga	actacttact	ctagcttccc	ggcaacaatt	16500
aatagactgg	atggaggcgg	ataaagtgtg	aggaccactt	ctgcgctcgg	cccttccggc	16560
tggctggttt	attgctgata	aatctggagc	cggtagcgtg	gggtctcggc	gtatcattgc	16620
agcactgggg	ccagatggta	agccctcccg	tatcgtagtt	atctacacga	cggggagtcg	16680
ggcaactatg	gatgaacgaa	atagacagat	cgctgagata	ggtgcctcac	tgattaagca	16740
ttggtaactg	tcagaccaag	tttactcata	tatactttag	attgatTTaa	aacttcattt	16800
ttaatttaaa	aggatctagg	tgaagatcct	ttttgataat	ctcatgacca	aaatccctta	16860
acgtgagttt	tcgttccact	gagcgtcaga	ccccgtagaa	aagatcaaa	gatcttcttg	16920
agatcccttt	tttctgcgcg	taatctgctg	cttgcaaaaca	aaaaaacac	cgctaccagc	16980
ggtggtttgt	ttgcgcggatc	aagagctacc	aactcttttt	ccgaaggtaa	ctgggttcag	17040
cagagcgcag	ataccaata	ctgtccttct	agtgtagccg	tagttaggcc	accacttcaa	17100
gaactctgta	gcaccgccta	catacctcgc	tctgctaata	ctgttaccag	tggctgctgc	17160
cagtggcgat	aagtctgtgc	ttaccgggtt	ggactcaaga	cgatagttac	cggataaggc	17220
gcagcggctg	ggctgaacgg	gggggttcgtg	cacacagccc	agcttggagc	gaacgacctt	17280
caccgaactg	agatacctac	agcgtgagct	atgagaaagc	gccacgcttc	cgaagggaga	17340
aaggcggaca	ggtatccggg	aagcggcagg	gtcggaaacg	gaga		17384

<210> 119

<211> 2814

<212> DNA

<213> Artificial Sequence

<220>

<223> pLITMUS38 Plasmid

<400> 119

gttaactacg	tcagggtggca	cttttcgggg	aaatgtgccc	ggaaccccta	tttgtttatt	60
tttctaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcggcc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgtttttgc	tcacccagaa	acgctggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagtctt	360
gctatgtggc	gcggtattat	cccggttgta	cgccgggcaa	gagcaactcg	gtcgccgcac	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tggcatgaca	gtaagagaat	tctgcagtgc	tgccataacc	atgagtata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccaaga	tgccctgtagc	aatggcaaca	acgttgcgca	aactattaac	720
tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	agggcgataa	780
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	840
tggagccggg	gagcgtgggt	ctcgcgggat	cattgcagca	ctggggccag	atggtaagcc	900
ctcccgtatc	gtagtatatc	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	ctttagattg	atttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatac	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttggt	aaatcagctc	attttttaac	caataggccg	aaatcggaac	aatcccttat	1200
aaatcaaaag	aatagcccga	gatagggttg	agtgtgtgtc	cagtttgga	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggcgcaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggg	cgaggtgccc	taaagcacta	1380
aatcggaaacc	ctaaagggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtc	1500
cgctgcgcgt	aaccaccaca	ccgcgcgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctagggtga	agatcctttt	tgataatctc	atgacaaaaa	tcccttaacg	tgagtttttcg	1620
ttccactgag	cgtcagaccc	cgtagaaaaa	atcaaaggat	cttcttgaga	tccctttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttgtttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaaggttaact	gcttcagcag	agcgcagata	1800
ccaaatactg	ttcttctagt	gtagccgtag	ttagggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaatacctg	ttaccagtg	ctgctgccag	tggcgataag	1920
tctgttctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	1980
tgaacggggg	gttcgtgcac	acagccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aaggagaaaa	ggcggacagg	2100

-89-

tatccggttaa	gcggcagggg	cggaacagga	gagcgcacga	gggagcttcc	agggggaaac	2160
gcctgggtatc	ttttatagtc	tgtcgggttt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgctcgt	cagggggg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttggctggc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggg	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcgggata	2400
acaatttcac	acaggaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gatatctgga	2520
tccacgaatt	cgctagcttc	ggccgtgacg	cgtctccgga	tgtacaggca	tgcgctcgacc	2580
ctctagtcaa	ggccttaagt	gagtcgtatt	acggactggc	cgctcgttta	caacgtcgtg	2640
actgggaaaa	ccctggcggt	acccaactta	atcgccctgc	agcacatccc	cctttcgcca	2700
gctggcgtaa	tagcgaagag	gcccgcaccg	atcgcccttc	ccaacagttg	cgcagcctga	2760
atggcggaatg	gcgcttcgct	tggtataataa	gcccgccttcg	gcgggctttt	tttt	2814

<210> 120

<211> 2847

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attB Plasmid

<400> 120

gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccoccta	tttgtttatt	60
tttctaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttcatcgaa	ctggatctca	acagcggtaa	300
gaccccttag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgcccgcat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tgccatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgctgtagc	aatggcaaca	acgttgcgca	aactattaac	720
tgccgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	780
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	840
tggagccggg	gagcgtgggt	ctcgcgggtat	catgtcagca	ctggggccag	atggtaagcc	900
ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	ctttagattg	atttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagatttgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtaa	1140
aatttttgtt	aaatcagctc	atttttttaac	caataggccg	aaatcgccaa	aatcccttat	1200
aaatcaaaaag	aatagcccga	gatagggttg	agtgttgttc	cagtttggaa	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggcgcaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aaccatcacc	taatttgggt	tttttgggtg	cgagggtgccg	taaaagcacta	1380
aatcggaacc	ctaaaggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgctgcgcgt	aaccaccaca	cccgccgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctagggtga	agatcctttt	tgataatctc	atgacaaaaa	tcccttaacg	tgagtttttcg	1620
ttccactgag	cgtcagaccc	cgtagaaaag	atcaaaggat	cttcttgaga	tcctttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaaacaaa	aaaccaccgc	taccagcggg	ggtttggtttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaaatactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaactcctg	ttaccagtgg	ctgctgccag	tgggcgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aagggagaaa	ggcggacagg	2100
tatccggttaa	gcggcagggg	cggaacagga	gagcgcacga	gggagcttcc	agggggaaac	2160
gcctgggtatc	tttatagtc	tgtcgggttt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgctcgt	cagggggg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttggctggc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggg	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcgggata	2400
acaatttcac	acaggaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gattgaagcc	2520
tgctttttta	tactaacttg	agcgaatct	ggatccacga	attcgctagc	ttcggccgtg	2580
acgcgtctcc	ggatgtacag	gcatgcgtcg	acccctctagt	caaggcctta	agtgagtcgt	2640
attacggact	ggccgtcggt	ttacaacgtc	gtgactggga	aaaccctggc	gttaccacaac	2700

-90-

ttaatgcgct	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	2760
ccgatcgccc	ttccccacag	ttgcgcgagcc	tgaatggcga	atggcgcttc	gcttggtaat	2820
aaagccccgct	tcggcgggct	ttttttt				2847

<210> 121

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA2 Plasmid

<400> 121

accatgaaaa	catttaacat	ttotcaacaa	gatctagaat	tagtagaagt	agcgacagag	60
aagattacaa	tgctttatga	ggataataaa	catcatgtgg	gagcggcaat	tcgtacgaaa	120
acaggagaaa	tcatttcggc	agtacatatt	gaagcgtata	taggacgagt	aactgtttgt	180
gcagaagcca	ttgcatggg	tagtgagtt	tcgaatggac	aaaaggattt	tgacacgatt	240
gtagctgtta	gacaccctta	ttctgacgaa	gtagatagaa	gtattcgagt	ggtaagtcct	300
tgtgggtatgt	gtaggaggtt	gatttcagac	tatgcaccag	attgttttgt	gttaatagaa	360
atgaatggca	agttagtcaa	aactacgatt	gaagaactca	ttccactcaa	atatacccg	420
aattaaaagt	ttaccatac	caagcttggc	tgctgcctga	ggctggacga	cctcgcgag	480
ttctaccggc	agtgcacatc	cgtcggcatc	caggaaacca	gcagcggcta	tcgcgcgcatc	540
catgcccccg	aactgcagga	gtggggaggg	acgatggccg	ctttggtccg	gatcttttgt	600
aagggaacctt	acttctgtgg	tgtagacataa	ttggacaaaac	tacctacaga	gatttaaagc	660
tctaaggtaa	atataaaatt	tttaagtgtta	taatgtgtta	aactactgat	tctaatttgt	720
tgtgtatttt	agattccaac	ctatggaact	gatgaatggg	agcagtgggtg	gaatgccttt	780
aatgagggaaa	acctgttttg	ctcagaagaa	atgccatcta	gtgatgatga	ggctactgct	840
gactctcaac	attctactcc	tccaaaaaag	aagagaaagg	tagaagacct	caaggacttt	900
ccttcagaat	tgctaagttt	tttgagtcac	gctgtgttta	gtaatagaac	tcttgcttgc	960
tttgctatatt	acaccacaaa	ggaaaaagct	gcactgctat	acaagaaaat	tatggaaaaa	1020
tattctgttaa	cctttataag	taggcataac	agttataatc	ataacatact	gttttttctt	1080
actccacaca	ggcatagagt	gtctgctatt	aataactatg	ctcaaaaatt	gtgtaccttt	1140
agcttttttaa	tttgttaaagg	gggttaataag	gaatatttga	tgtatagtgc	cttgactaga	1200
gatcataatc	agccatacca	catttgtaga	ggttttactt	gctttaaaaa	acctcccaca	1260
cctccccctg	aacctgaaac	ataaaatgaa	tgcaatttgt	gttgtaaact	tgttttattgc	1320
agcttataat	ggttacaaat	aaagcaatag	catcacaaat	ttcacaaaata	aagatccaga	1380
tttcgctcaa	gtagtatata	aaaagcaggc	ttcaatcctg	cagagaagct	tggcgccagc	1440
cggcttcaat	tgcacgggccc	ccactagtga	gtcgtattac	gtagcttggc	gtaatcatgg	1500
tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	catacgagcc	1560
ggaagcataa	agtgtaaagc	ctgggggtgcc	taattgagtga	gctaactcac	attacatgtg	1620
agcaaaaaggc	cagcaaaaagg	ccaggaaccg	taaaaaaggcc	gcgttgctgg	cgtttttcca	1680
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccga	gggtggcgaaa	1740
ccgcagagga	ctataaagat	accaggcggt	tcccctcgga	agctccctcg	tgcgctctcc	1800
tggtccgacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	1860
gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	taggtcgttc	gctccaagct	1920
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	1980
tcttgagtc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	2040
gatttagcaga	gcgaggtatg	taggcgggtgc	tacagagttc	ttgaagtggg	ggcctaacta	2100
cggctacact	agaagaacag	tatttgggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	2160
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggttagcg	gtgggtttttt	2220
tgtttgcaag	cagcagatta	cgcgagaaa	aaaaggatct	caagaagatc	ctttgatctt	2280
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggaattt	tggtcatgag	2340
attatcaaaa	aggatcttca	cctagatcct	tttacgcgcc	ctgtagcggc	gcattaagcg	2400
cggcggtgtg	gggtgttacg	cgcagcgtga	ccgctacact	tgccagcgcc	ctagcgcccc	2460
ctcctttcgc	tttcttcctt	tcctttctcg	ccacggttcg	tttcccgcgc	aagctctaaa	2520
tcggggggtc	cctttagggt	tccgatttag	tgctttacgg	cacctcgacc	ccaaaaaact	2580
tgatttgggt	gatggttcac	gtagtggggc	atcgccctga	tagacggttt	ttcgcccttt	2640
gacgttggag	tccacgtttc	ttaatagtgg	actcttggtc	caaactggaa	caacactcaa	2700
ccctatctcg	ggctattctt	ttgatttata	agggattttg	ccgatttcgg	cctattgggt	2760
aaaaaatgag	ctgatttaac	aaaaatttaa	cgcgaatttt	aacaaaatat	taacgtttac	2820
aattttaaata	tttgcttata	caatcttcct	gtttttgggg	cttttctgat	tatcaaccgg	2880
ggtaaatcaa	cttaagtagt	atatgagtaa	atttggtctg	acagttacca	atgcttaatc	2940
agtgaggcac	ctatctcagc	gatctgtcta	tttcgttcat	ccatagttgc	ctgactcccc	3000
gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtgc	tgcaatgata	3060
ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccgggaag	3120
gccgagcgca	gaagtgggtc	tgcaacttta	tcgcgctcca	tccagtctat	taattgttgc	3180
cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	3240

-91-

acaggcatcg	tgggtgcacg	ctcgtcggtt	ggtatggctt	cattcagctc	cgggtcccaa	3300
cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	aagcgggttag	ctccttcggg	3360
cctccgatcg	ttgtcagaag	taagttggcc	gcagtgttat	cactcatggg	tatggcagca	3420
ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	tgggtgagta	3480
tcaaccaagt	cattctgaga	atagtgatg	cggcgaccga	gttgctcttg	cccgccgcta	3540
acacgggata	ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	tggagaacgt	3600
tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	3660
actcgtgcac	ccaactgatc	ttcagcatct	tttactttca	ccagcgtttc	tgggtgagca	3720
aaaacaggaa	ggcaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	3780
ctcatactct	tcctttttca	atattattga	agcattttatc	agggttattg	tctcatgagc	3840
ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	gggttccgcg	cacatttccc	3900
cgaaaagtgc	cacctgacgt	agttaacaaa	aaaaagcccg	ccgaagcggg	ctttattacc	3960
aagcgaagcg	ccattcgcca	ttcaggctgc	gcaactgttg	ggaaggcgca	tcgggtcggg	4020
cctcttcgct	attacgccag	ttggcgaaag	ggggatgtgc	tgcaaggcga	tttaagttggg	4080
taacgccagg	gttttcccag	tcacgacgtt	gtaaaacgac	ggccagtcgg	taatacgact	4140
cacttaaggc	cttgactaga	gggtcgacgc	atgcctgtac	atccggagac	gcgtcacggc	4200
cgaagctagc	gaattcgtgg	atc				4223

<210> 122

<211> 2686

<212> DNA

<213> Artificial Sequence

<220>

<223> pUC18 Plasmid

<400> 122

tcgcgcgttt	cgggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aactgttggg	aaggcgcatc	ggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaaaagg	ggatgtgctg	caaggcgatt	aagttgggtg	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgccaa	gcttgcatgc	ctgcaggctc	420
actctagagg	atccccgggt	accgagctcg	aattcgtaat	catggtcata	gctgtttcct	480
gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	540
aaagcctggg	gtgcctaatt	agttagctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	600
gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	gaatcgccca	acgcgcgggg	660
agaggcggtt	tcggtattgg	gcgctcttcc	gcttcctcgc	tcactgactc	gctgcgctcg	720
gtcgttcggc	tgcggcgagc	ggtatcagct	cactcaaaag	cggtaatagc	gttatccaca	780
gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	ggccaggaac	840
cgtaaaaagg	ccgcgttgct	ggcggttttc	cataggctcc	gccccctga	cgagcatcac	900
aaaaatcgac	gctcaagtca	gaggtggcga	aaccgcagag	gactataaag	ataccaggcg	960
tttcccctcg	gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	1020
ctgtccgcct	ttctcccttc	gggaagcgtg	ccgctttctc	atagctcacg	ctgtagggtat	1080
ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	1140
cccgaccgct	gcgccttatc	cggtaactat	cgtcttgagt	ccaaccgggt	aagacacgac	1200
ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	1260
gtacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	1320
atctgcgctc	tgtgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	1380
aaacaaacca	ccgctggtag	cgggtggttt	tttgtttgca	agcagcagat	tcagcgcaga	1440
aaaaaaaggat	ctcaagaaga	tccttttgatc	ttttctacgg	ggtctgacgc	tcagtggaaac	1500
gaaaactcac	gttaagggat	tttggctcatg	agattatcaa	aaaggatcct	cacctagatc	1560
cttttaaat	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttggtct	1620
gacagttacc	aatgcttaat	cagtgaaggca	cctatctcag	cgatctgtct	atttcgttca	1680
tccatagtgt	cctgactccc	cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	1740
ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	1800
ataaaccagc	cagccggaag	ggccgagcgc	agaagtgggtc	ctgcaacttt	atccgcctcc	1860
atccagtcta	tttaattgttg	ccgggaagct	agagtaagta	gttcgcccagt	taatagtttg	1920
cgcaacgttg	ttgccattgc	tacaggcatc	gtgggtgcac	gctcgtcgtt	tgggtatggct	1980
tcattcagct	ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	2040
aaagcggtta	gctccttcgg	tcctccgac	gttgtcagaa	gtaagttggc	cgcagtggtta	2100
tcactcatgg	ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	2160
ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	aatagtgat	gcggcgaccg	2220
agttgctctt	gccccgcgtc	aataccggat	aataccgcgc	cacatagcag	aactttaaaa	2280
gtgctcatca	tttgaaaacg	ttcttcgggg	cgaaaactct	caaggatcct	accgctgttg	2340
agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	2400

-92-

accagcggtt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataagg	2460
gcgacacgga	aatggttgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	2520
cagggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	2580
ggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	tctaagaaac	cattattatc	2640
atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc		2686

<210> 123

<211> 8521

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFPattB(6xHS4)2 Plasmid

<400> 123

tacggggcgg	gggatccact	agttattaat	agtaaatcaat	tacgggggtca	ttagttcata	60
gcccatatat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgt	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccc	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcac	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	gggggcgcgc	gccaggcggg	gcggggcggg	gcgagggggc	540
ggggcggggc	aggcggagag	gtgcggcggc	agccaatcag	agcggcgcg	tccgaaagtt	600
tccttttatg	gcgaggcggc	ggcgcgcgcg	gccctataaa	aagcgaagcg	cgcggcgggc	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgtccgc	gccgcctcgc	gccgcccggc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctctccc	780
gggctgtaac	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgctgaaag	840
ccttaaaggg	ctccgggagg	gcccttttgt	cgggggggag	cggtcggggg	ggtgctgctg	900
tgtgtgtgtg	cgtggggagg	gccgcgtgcg	gccgcgctg	cccggcggct	gtgagcgtg	960
cgggtcccg	ggcgggcctt	gtgcgctcgc	cgtgtgcgcg	aggggagcgc	ggccgggggc	1020
gggtccccgc	ggtgcggggg	ggctgcgagg	gggaacaaag	ctgcgtgcgg	ggtgtgtgctg	1080
tgggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccgctc	tcgggtgcgg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	ccgtgccggg	cgggggggtg	cggcaggtgg	gggtgccggg	cggggcgggg	1260
ccgcctcggg	ccgggggagg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggtc	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcggtcgag	agggcgcaag	1380
gacttccttt	gtcccaaatc	tggcggagcc	gaaatctggg	aggcgccgcc	gcacccccct	1440
tagcggggcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctcgct	ctggggggga	cggggcaggg	cgggggttcg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctacacg	ctcctgggca	1680
acgtgctggg	tgttgtgtgt	tctcatcatt	ttggcaaaag	attcgccacc	atggtgagca	1740
agggcgagga	gctgttcacc	ggggtggtgc	ccatcctggt	cgagctggac	ggcgacgtaa	1800
acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgacct	cggcgtgcag	tgcttcagcc	gctaccccga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gccgaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcgcg	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatgggtc	ctgctggagt	2400
tcgtgaccgc	cgccgggata	actctcggca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtg	tggctggtgt	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagcccttgg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tgttggattt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcatttaaaa	catcagaatg	2700
agtaatttgg	ttagagtttg	gcaacatatg	ccatatgctg	gctgcccata	acaaagggtg	2760
ctataaagag	gtcatcagta	tatgaacacg	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgaggtta	gatttttttt	atattttgtt	tttgtgtatt	tttttcttta	2880
acatccctaa	aattttcctt	acatgtttta	ctagccagat	ttttcctcct	ctcctgacta	2940
ctcccagtc	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gcccagcttt	3000
ggcgtaatca	tggtcatagc	tgtttctctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaattgag	tgagctaact	3120

-93-

cacattaatt	gcgttgcgct	caactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcgat	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgccc	3240
ctaactccgc	ccagttccgc	ccattctccg	cccatggct	gactaatttt	ttttatttat	3300
gcagaggccg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggctagt	ggatcccccg	ccccgtatcc	cccaggtgtc	tgcaggctca	aagagcagcg	3420
agaagcgttc	agaggaaagc	gatcccgtgc	caccttcccc	gtgcccgggc	tgtcccgcga	3480
cgtgcccggc	tcggggatgc	ggggggagcg	ccggaccgga	gcgagagccc	gggaggctcg	3540
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	3600
gtcccgcgtg	gcggtatccg	ggccccgtat	cccccagggt	tctgcaggct	caaagagcag	3660
cgagaagcgt	tcagaggaaa	gcgatcccg	gccaccttcc	ccgtgcccgg	gctgtccccg	3720
cacgtgccc	gctcggggat	gcggggggag	cgccggaccg	gagcggagcc	ccgggcggct	3780
cgctgctgcc	ccctagcggg	ggaggggacgt	aattacatcc	ctgggggctt	tggggggggg	3840
ctgtccccgt	gagcggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	3900
agcgagaagc	gttcagagga	aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	3960
cgcacgctgc	cggctcgggg	atgcgggggg	agcgccggac	cggagcggag	ccccgggcgg	4020
ctcgtctgct	ccccctagcg	gggggagggg	gtaattacat	ccctgggggc	tttggggggg	4080
ggctgtcccc	gtgagcggat	ccgcggccccc	gtatccccc	ggtgtctgca	ggctcaaaga	4140
gcagcgagaa	gcgttcagag	gaaagcgatc	ccgtgccacc	ttccccgtgc	ccgggctgtc	4200
cccgcacgct	gcccggctcg	ggatgcgggg	ggagcgccgg	accggagcgg	agccccgggc	4260
ggctcgctgc	tgcctccctag	cggggggagg	acgttaattac	atccctgggg	gctttggggg	4320
ggggctgtcc	ccgtgagcgg	atccgcggcc	ccgtatcccc	cagggtgtct	caggctcaaa	4380
gagcagcgag	aagcgttcag	aggaaagcga	tcccgtgcc	ccttccccgt	gcccgggctg	4440
tccccgcacg	ctgcgggctc	ggggatgcgg	ggggagcgcc	ggaccggagc	ggagccccgg	4500
gcggctcgct	gctgccccct	agcgggggag	ggagcgtaat	acatccctgg	gggctttggg	4560
gggggctgtg	ccccgtgagc	ggatcccgcg	ccccgtatcc	cccagggtgt	tgcaggctca	4620
aagagcagcg	agaagcgttc	agaggaaagc	gatcccgtgc	caccttcccc	gtgcccgggc	4680
tgtccccgca	cgtgcccggc	tcggggatgc	ggggggagcg	ccggaccgga	cgggagcccc	4740
ggcgcgctcg	ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	4800
gggggggggct	gtccccgtga	gcggatccgc	ggggctgcag	gaattcgatt	gaagcctgct	4860
tttttatact	aacttgagcg	aaatcaagtc	cctaggcttt	tgcaaaaagc	taacttggtt	4920
attgcagctt	aaatgggta	caaataaagc	aatagcatca	caaatttcac	aaataaagca	4980
ttttttcac	tgcattctag	ttgtggtttg	tccaaactca	tcaatgtatc	ttatcatgtc	5040
tggatccgct	gcattaatga	atcgcccaac	gcgcggggag	aggcggtttg	cgtattgggg	5100
gctcttccgc	ttcctcgctc	actgactcgc	tgcgtcgggt	cgttcggctg	cggcgagcgg	5160
tatcagctca	ctcaaaggcg	gtaatacggg	ttccacaga	atcaggggat	aacgcaggaa	5220
agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	5280
cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtcaag	5340
ggtgcgaaaa	ccgcacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	5400
tgcgtcttcc	tgttcgcacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	5460
gaagcgtggc	gctttctcaa	tgtctacgct	gtaggtatct	cagttcgggt	taggtcgttc	5520
gctccaagct	gggtgtgtgt	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	5580
gtaactatcg	tcttgagtcg	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	5640
ctggtaacag	gattagcaga	gcgaggatag	taggcgggtg	tacagagttc	ttgaagtggg	5700
ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	5760
ttactctcgg	aaaagagttt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	5820
gtgggttttt	tgtttgcaag	cagcagatta	cgcgcaaaaa	aaaaggatct	caagaagatc	5880
ctttgatctt	ttctacgggg	tctgacgctc	agtggaaagc	aaactcacgt	taagggattt	5940
tggatcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aaatgaagtt	6000
ttaaatcaat	ctaaagtata	tatgagtaaa	cttggctcga	cagttaccaa	tgcttaatca	6060
gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcac	catagttgcc	tgactccccg	6120
tcgtgtagat	aactacgata	cgggaggggt	taccatctgg	ccccagtgct	gcaatgatac	6180
cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	6240
ccgagcgcag	aagtgggtcct	gcaactttat	ccgcctccat	ccagttctat	aattggtgcc	6300
gggaagctag	agtaagtagt	tcgcccagtt	atagtttgcg	caacgttggt	gccattgcta	6360
caggcatcgt	gggtgcacgc	tcgtcggttg	gtatggcttc	attcagctcc	ggttcccaac	6420
gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	agcgggttag	tccttcgggt	6480
ctccgatcgt	tgtcagaagt	aagtggcccg	cagtggttat	actcatgggt	atggcagcac	6540
tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	6600
caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	6660
tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacggt	6720
cttcggggcg	aaaactctca	aggatctttc	cgctgttgag	atccagttcg	atgtaaccca	6780
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	6840
aaacagggaag	gcataaatg	gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	6900
tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	6960
gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	7020
gaaaagtgcc	acgtggtcga	cgttatcgat	aagcttgata	tcgaattcct	gcagccccgc	7080
ggatccgctc	acggggacag	ccccccccc	aagccccag	ggatgtaatt	acgtccctcc	7140

-94-

```

ccccgctaggg  ggcagcagcg  agccgcccgg  gggtccgctc  cgggtccggcg  cccccccgc  7200
atccccgagc  cggcagcgtg  cggggacagc  ccggggcacgg  ggaagggtggc  acgggatcgc  7260
tttccctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  atacggggcc  7320
gcggtatccgc  tcacggggac  agcccccccc  caaagccccc  agggatgtaa  ttacgtccct  7380
cccccgctag  gggggcagcag  cgagccgccc  ggggctccgc  tccggtccgg  cgctcccccc  7440
gcatccccga  gccggcagcg  tgcggggaca  gcccgggcac  ggggaagggtg  gcacgggatc  7500
gctttcctct  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  ggatacgggg  7560
ccgcggatcc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  aattacgtcc  7620
ctcccccgct  agggggcagc  agcagccgc  ccgggggtcc  gctccggtcc  ggcgctcccc  7680
ccgcatcccc  gagccggcag  cgtgcgggga  cagcccgggc  acggggaagg  tggcacggga  7740
tcgctttcct  ctgaacgctt  ctgctgctc  tttgagcctg  cagacacctg  ggggatacgg  7800
ggcgcggtat  ccgctcacgg  ggacagcccc  ccccacaaag  cccagggat  gtaattacgt  7860
ccctcccccg  ctagggggca  gcagcgagcc  ccgcccgggt  ccgctccggt  ccggcgctcc  7920
ccccgcatcc  ccgagccggc  agcgtgcggg  gacagcccg  gcacggggaa  ggtggcacgg  7980
gatcgctttc  ctctgaacgc  ttctcgctgc  tctttgagcc  tgcagacacc  tgggggatcc  8040
ggggccgcg  atccgctcac  ggggacagcc  cccccccaaa  gccccagg  atgtaattac  8100
gtccctcccc  cgctaggggg  cagcagcgag  ccgcccgggg  ctccgctccg  gtccggcgct  8160
ccccccgcat  ccccgagccg  gcagcgtg  gggacagccc  gggcacgggg  aagggtggc  8220
gggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  8280
acggggccgc  ggatccgctc  acggggacag  cccccccca  aagccccag  ggatgtaatt  8340
acgtccctcc  ccgctaggg  ggcagcagcg  agccgcccgg  gggtccgctc  cgggtccggc  8400
ctccccccgc  atccccgagc  cggcagcgtg  cggggacagc  ccggggcacgg  ggaagggtggc  8460
acgggatcgc  tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  8520
a

```

<210> 124

<211> 8851

<212> DNA

<213> Artificial Sequence

<220>

<223> p18EPOcDNA Plasmid

<400> 124

```

cagttgccgg  ccgggtcgcg  cagggcgaac  tcccgcccc  acggtgctc  gccgatctcg  60
gtcatggccg  gcccgaggc  gtcccgaag  ttctgtgaca  cgacctccga  ccactcggcg  120
tacagctcgt  ccaggccgg  caccacacc  caggccaggg  tggtgtccgg  caccacctgg  180
tcctggaccg  cgctgatgaa  cagggtcacg  tcgtcccgga  ccacaccggc  gaagtcgtcc  240
tccacgaagt  cccgggagaa  cccgagccgg  tcggtccaga  actcgaccgc  tccggcgacg  300
tcgcgcgcgg  tgagcaccgg  aacggcactg  gtcaacttgg  ccatggatcc  agatttcgct  360
caagtttagta  taaaaaagca  ggcttcaatc  ctgcagagaa  gcttgatata  gaattcctgc  420
agcccccgcg  atccgctcac  ggggacagcc  cccccccaaa  gccccagg  atgtaattac  480
gtccctcccc  cgctaggggg  cagcagcgag  ccgcccgggg  ctccgctccg  gtccggcgct  540
ccccccgcat  ccccgagccg  gcagcgtg  gggacagccc  gggcacgggg  aagggtggc  600
gggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  660
acggggccgc  ggatccgctc  acggggacag  cccccccca  aagccccag  ggatgtaatt  720
acgtccctcc  ccgctaggg  ggcagcagcg  agccgcccgg  gggtccgctc  cgggtccggc  780
ctccccccgc  atccccgagc  cggcagcgtg  cggggacagc  ccggggcacgg  ggaagggtggc  840
acgggatcgc  tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  900
atacggggcc  gcggtatccg  tcacggggac  agcccccccc  caaagccccc  agggatgtaa  960
ttacgtccct  ccccgctag  ggggcagcag  cgagccgccc  ggggctccgc  tccggtccgg  1020
cgctcccccc  gcatccccga  gccggcagcg  tgcggggaca  gcccgggcac  ggggaagggtg  1080
gcacgggatc  gctttcctct  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  1140
ggatacgggg  ccgcggatcc  gctcacgggg  acagccccc  cccaaagccc  ccagggatgt  1200
aattacgtcc  ctcccccgct  agggggcagc  agcgagccgc  ccgggggtcc  gctccggtcc  1260
ggcgctcccc  ccgcatcccc  gagccggcag  cgtgcgggga  cagcccgggc  acggggaagg  1320
tggcacggga  tcgctttcct  ctgaacgctt  ctgctgctc  tttgagcctg  cagacacctg  1380
ggggatacgg  ggcccggtat  ccccgccgg  ctagggggca  ggacagcccc  ccccacaaag  1440
gtaattacgt  ccctcccccg  ccgagccggc  agcgtgcggg  gcagcgagcc  gcccggggt  1500
ccggcgctcc  ccccgcatcc  ctctgaacgc  ttctcgctgc  ggggacagcc  ccccggggt  1560
gggtggcacgg  ggggcggcg  atccgctcac  cgggacagcc  ccccggggt  ccgctccggt  1620
tgggggatac  gtccctcccc  cgctaggggg  gcagcgtg  gggacagccc  tctttgagcc  1680
atgtaattac  gtcggcgct  ccccgcgat  ccccgagcc  gctctttgag  gggcacgggg  1740
gtccggcgct  aaggtggcac  cctgggggat  tagttcatag  cccatatatg  gagttccg  1800

```

gctgaccgcc	caacgacccc	cgccattga	cgtaataat	gacgtatgtt	cccatagtaa	2040
cgcaaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgtctc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggcggggg	2400
cgaggggcg	ggcggggcca	ggcgagagg	tgcgggcgca	gccaatcaga	gcggcgcgct	2460
ccgaaagt	ccttttatgg	cgaggcggg	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtcgctg	cgttgccttc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccggcc	cggtcttgac	tgaccgctt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgcttggt	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tccgggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	cccgcgctgc	ccggcgctgc	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gccggggcg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgctgcgggg	2940
gtgtgtgcgt	gtgggggtga	gcagggggtg	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgacacccc	ctccccgagt	tgctgagcac	ggcccggctt	cggtgtcggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgcggggc	ggggggtggc	ggcaggtggg	ggtgcggggc	3120
ggggcggggg	cgctcgggc	cggggagggc	tcgggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcgggcg	gtcagggcgc	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcctttg	tcccaaatct	ggcgagcccg	aaatctggga	ggcgccggccg	3300
cacccctct	agcggggcgc	ggcgaagcgg	tgccggcgccg	gcaggaagga	aatgggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcggcgct	ccccctctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggtgcctt	cgggggggac	ggggcagggc	ggggttcggc	ttctggcggtg	3480
tgaccggcg	ctctagaatg	ggggtgcacg	aatgtcctgc	ctggctgtgg	cttctcctgt	3540
ccctgctgtc	gctccctctg	ggcctccag	tcctggcgcg	cccaccacgc	ctcatctgtg	3600
acagccgagt	ctggagagg	tacctcttgg	aggccaagga	ggccgagaat	atcacgacgg	3660
gctgtgctga	acactgcagc	ttgaatgaga	atatcactgt	cccagacacc	aaagttaatt	3720
tctatgcctg	gaagaggatg	gaggtcgggc	agcaggccctg	agaagtctgg	cagggcctgg	3780
ccctgctgtc	gtgcgtcgct	ctgcggggcc	agccctgtgt	ggtcaactct	tcccagccgt	3840
gggagccct	gcagctgcat	gtggataaag	ccgtcagtgg	ccttcgcagc	ctcaccactc	3900
tgcttcgggc	tctgggagcc	cagaaggaag	ccatctcccc	tccagatgcg	gcctcagctg	3960
ctccactccg	aacaatcact	gctgacactt	tccgcaact	cctccgagtc	tactccaatt	4020
tcctccgggg	aaagctgaag	ctgtacacag	gggaggcctg	caggacaggg	gacagatgac	4080
gtacaagtaa	gaattcactc	ctcaggtgca	ggctgcctat	cagaaggtgg	tggtcggtgt	4140
ggccaatgcc	ctggctcaca	aataccactg	agatcttttt	ccctctgcca	aaaattatgg	4200
ggacatcatg	aagccccttg	agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	4260
tgcaatagt	tgttggaatt	ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	4320
tcaattaaaa	ctcagaaatg	agtatttgg	ttagagtgtg	gcaacatatg	ccatatgctg	4380
gctgcaatga	acaaaggtgg	ctataaagag	gtcatcagta	tatgaaacag	ccccctag	4440
tccattcctt	attccataga	aaagccttga	cttgaggtta	gatttttttt	atattttgtt	4500
ttgtgttatt	tttttcttta	acatccctaa	aattttctct	acatgtttta	ctagccagat	4560
ttttcctcct	ctcctgacta	ctcccagtc	tagctgtccc	tcttctctta	tgaagatccc	4620
tcgacctgca	gcccagctt	gcatgcctgc	aggtcgactc	tagtggatcc	ccccggcgt	4680
atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	aagcgatccc	4740
gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgctgc	cggtcggggg	atgcgggggg	4800
agcgcgggac	cggaagcgag	ccccgggcgg	ctcgctgctg	ccccctagcg	ggggaggggc	4860
gtaattacat	ccctgggggc	tttggggggg	ggctgtcccc	gtgagcggat	ccgcggcccc	4920
gtatccccca	gggtgtctgca	ggctcaaaga	gcagcgagaa	gcgttcagag	gaaagcgatc	4980
ccgtgccacc	ttccccgtgc	ccgggctgtc	cccgacgct	gcccggctcgg	ggatgcgggg	5040
ggagcgccgg	accggagcgg	agccccgggc	ggctcgctgc	tgccccctag	cgggggaggg	5100
acgtaat	atccctgggg	gctttggggg	gggctgtccc	ccgtgagcgg	atccgcggcc	5160
ccgtatcccc	caggtgtctg	caggtc	gagcagcgag	aagcgttcag	aggaaagcga	5220
tcccgctgca	ccttccccgt	gcccgggctg	tccccgcacg	ctgcggctc	ggggatgcgg	5280
ggggagcgcc	ggaccggagc	ggagccccgg	gcggctcgct	gctgccccct	agcgggggag	5340
ggacgttaatt	acatccctgg	gggctttggg	ggggggctgt	ccccgtgagc	ggatccgcgg	5400
ccccgtatcc	cccaggtgtc	tgagggtca	aagagcagcg	agaagcgttc	agaggaaagc	5460
gatccccgtg	caccttcccc	gtgcccgggc	tgtccccgca	cgctgcccgc	tcgggggatgc	5520
ggggggagcg	ccggaccgga	gcggagcccc	gggcggctcg	ctgctgcccc	ctagcggggg	5580
agggacgtaa	ttacatccct	gggggctttg	ggggggggct	gtccccgtga	gcggatccgc	5640
gcccccgat	ccccaggtg	tctgcagggt	caaagagcag	cgagaagcgt	tcagaggaaa	5700
gcgatcccg	gccaccttcc	ccgtgcccgg	gctgtccccg	cacgctgccc	gctcggggat	5760
gcgggggggag	cgccggaccg	gagcggagcc	ccggcgggct	cgctgctgcc	ccctagcggg	5820
ggagggagcgt	aattacatcc	ctgggggctt	tggggggggg	ctgtccccgt	gagcggatcc	5880
gcgggcccgt	atccccagg	gtgtgcagg	ctcaagagc	agcgagaagc	gttcagagga	5940
aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgctgc	cggtcggggg	6000

-96-

```

atgcggggggg agcgccgggac cggagcgggag ccccgggggg ctcgctgctg cccctagcg 6060
ggggaggggac gtaattacat ccctggggggg ttgggggggg ggctgtcccc gtgagcgggat 6120
ccgcgggggct gcaggaattc gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat 6180
ccgctcacaa ttccacacaa catacagagcc ggagcataaa agtgtaaagc ctgggggtgcc 6240
taatgagtga gctaactcac attaatgtcg ttgcgctcac tgcccgcgtt ccagtcggga 6300
aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg cgggtttgcgt 6360
attgggcgct cttccgcttc ctogctcact gactcgctgc gctcggtcgt tcggctgcgg 6420
cgagcgggat cagctcactc aaaggcggta atacggttat ccacagaatc aggggataaac 6480
gcaggaaaga acatgtgagc aaaaggccag cctgacgagc ggaaccgtaa aaaggccgcg 6540
ttgctggcgt ttttccatag gctccgcccc taaagataacc aggcgtttcc ccctggaagc 6600
agtccagagt ggcgaaaccc tccgacctg ccgcttaccg gatacctgtc cgctttctc 6720
tccctcgtgc gctcctcgtg ttctcatagc tcacgctgta ggtatctcag ttcgggtgtag 6780
ccttcgggaa gcgtggcgct ctgtgtgcac gaaccccccg ttcagcccg cgcgtgcgc 6840
gtcgttcgct ccaagctggg tgagtccaac ccgtaagac acgacttacc gccactggca 6900
ttatccggta actatcgtct tagcagagcg aggtatgtag gcggtgctac agagttcttg 6960
gcagccactg gtaacaggat ctacactaga aggcagtat ttggtatctg cgctctgctg 7020
aagtgggtgg ctaactacgg aagagttggg cagatctgat ccggcaaaaa aaccaccgct 7080
aagccagtta ccttcggaaa ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa 7140
ggtagcgggtg gtttttttgt tacggggctc gagcgtcagt ggaacgaaaa ctcacgttaa 7200
gaagatcctt gggatttttg atcaaaaagg atcttcacct agatcctttt aaattaaaaa 7260
tgaagtttta aatcaatcta aagtatatat ctcagcgatc gagtaaactt ggtctgacag ttaccaatgc 7320
ttaatcagtg aggcacctat tacgatacgg gagggtttac catctggccc cagtgtgca 7380
ctccccgtcg tgtagataac ctacccacg ctacccggct cagcaataaa ccagccagcc 7440
atgataccgc agcgacagaag tggctcctgca actttatccg cctccatcca gtctattaat 7500
ggaagggccg aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc 7560
tgttgccggg atgctacag gcatcggtgt tacatgatcc ccatgttgtt gtttgccgag 7620
attgctacag caaggcgagt cgatcgttgt tggcccgtag gatgctttc gatgctttc 7680
tcccaacgat tccggtcctc gactcgttgt cagaagtaag ttggcccgag gaccgagttg ctcttgccc 7740
gcagcactgc ataattctct ctgagaaatg agcagaactt taaaagtgtc catcattgga 7800
gagtactcaa ccaagtcatt ctcgctcatg atctccgtaa gatgctttc tgtgactggt 7860
gcgtcaatac gggataatac cgcgccacat agcagaactt taaagtgtc ctcttgccc 7920
aaacgttctt cggggcgaaa actctcaagg atcttaccgc taaagtgtc catcattgga 7980
taacccactc gtgcacccaa ctgatcttca gcactttta atcttaccg cagttcgatg 8040
tgagcaaaaa caggaaggca aaatgccgca tattgaagca aaataggggt agcttcgatg 8100
tgaatactca tactcttcct ttttcaatat aaaaagggaa taagggcgac acggaaatgt 8160
atgagcggat acatatttga atgtatctag tgacgtagtt ttttcccgaa agcccgcgga agcgggcttt 8220
tttcccgaa aagtgccacc tgcgctatta cccagtcac cgttgaggaa gggcgatcgg 8280
attaccaagc gaagcgccat cgccagctgg cgaaaggggg atgtgctgca aggcgattaa 8340
tgccggcctc ttcgctatta tccagtcac caggttgtaa aacgacggcc agtcctgtaa 8400
gttgggtaac gtcagggttt actagagggt caggttgtaa aacgacggcc agtcctgtaa 8460
acgactcact taaggccttg gacaaaccac aactagaatg cagcggtata cagacatgat aagatacatt 8520
gatgagtttg gacaaaccac aactagaatg cagcggtata cagacatgat aagatacatt 8580
tgtgatgcta tgcgtttatt tgtaaccatt ccccgcgctg gaggtatcat cagccggcgt 8640
gaagaactcc agcatgagat ccccgcgctg gaggtatcat cagccggcgt 8700
gattccgaag cccaaccttt catagaaggc ggcggtggaa tcgaaatctc cccggaaaac 8760
tcagtcctgc tectcgccca cgaagtgcac gggcggtggg ctcgaaatctc gtagcacgtg 8820
8851

```

<210> 125

<211> 10474

<212> DNA

<213> Artificial Sequence

<220>

<223> p18genEPO Plasmid

<400> 125

```

cagttggcgg cggggtcgcg cagggcgaac tcccgcccc acggctgctc gccgatctcg 60
gtcatggcgg gcccgaggc gtcccggaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggccgcy caccacaccc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccga ccacaccgg ccaggcgacg 240
tccacgaagt cccggagaa cccgagccgg tccgtccaga actcgaccgc tccggcgacg 300
tcgcgcgcg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatata gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc ccccccaggg gcccccagg gtcgggcgct 480
gtccctcccc cgctaggggg cagcagcgag ccgccccggg ctccgctccg gtccggcgct 540

```

-97-

ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	aagggtggcac	600
gggatoctct	tctctctgaac	gcttctctcgt	gctcttttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgctaggg	ggcagcagcg	agccgcccgg	ggctccgctc	cgggtccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggcacgg	ggaagggtggc	840
acgggatccg	tttctctctga	acgcttctcg	ctgctctttg	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggac	agcccccccc	caaagcccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagccgccc	ggggctccgc	tccggtcccg	1020
cgctcccccc	gcattccccga	gccggcagcg	tgccggggaca	gccccgggac	gggggaagggtg	1080
gcacgggatc	gctttctctc	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	cccgcgatcc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcagagccg	ccggggctcc	gctccgggtcc	1260
ggcgctcccc	ccgcatcccc	gagccggcag	cgtgcggggg	cagccccggg	acggggaagg	1320
tggcacggga	tcgcttttct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggccgcggat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440
gtaattacgt	ccctcccccg	ctagggggca	gcagcagagc	gccccggggc	ccgctccggg	1500
ccggcgctcc	ccccgcaccc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
gggtggcacg	gatcgctttc	ctctgaacgc	ttctcgctgc	tctttgagcc	tgacagaccc	1620
tgggggatac	ggggcccgcg	atccgctcac	ggggacagcc	cccccccaaa	gcccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcag	ccgccccggg	ctccgctccg	1740
gtccggcgct	ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aagggtggcac	gggatcgctt	tctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgccctg	1980
gctgaccgcc	caacgacccc	cgccattga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacgta	2160
aatggccccg	ctggcattat	gcccagtaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctcccccc	ctccccaccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cagtgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcgggggcg	ggcgggaggg	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagtgt	ccttttatgg	cgaggcgggc	gccccggggc	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cgttgccctc	gccccgtgcc	ccgctccgcg	ccgctccgcg	2580
ccgcccgcgc	cggtctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctctcccg	ggctgtaatt	agcgtttgg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaaaggg	tccgggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgctg	gtggggagcg	ccgcgtgctg	cccgcgctgc	ccggcgctgc	2820
tgagcgctgc	ggggcgcgcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gccccggggc	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggg	tgctgcccgg	2940
gtgtgtgcgt	gggggggtga	gcaggggggt	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctcccagagt	tgctgagcac	ggcccggtt	cggtgctggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgccgggc	gggggggtgg	ggcagggtgg	ggtgccgggc	3120
ggggcggggg	cgctctgggg	cggggagggc	tccgggggag	ggcgcgggcg	ccccggagcg	3180
ccggcgcgct	tcagagcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttctcttg	tcccaaactc	ggcgggagcc	aaatctggga	ggcgcccgcc	3300
cacccccctc	agcggcgcg	ggcgaaagcg	tgccggcgcc	gcaggaaagg	aatggggcg	3360
gagggccttc	gtgcgtcgcc	gcgcccgcgt	ccccttctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggtgcctt	cgggggggac	ggggcgaggc	gggttcggc	ttctggcggt	3480
tgaccggcg	ctctagatgc	atgctcgagc	ggccgcaggt	gtgatggata	tctgcagaat	3540
tcgccccttc	tagaatgggg	gtgcacggtg	agtaactcg	ggctgggcgc	ccccgcccgc	3600
ccgggtccct	gtttgagcgg	ggatttagcg	ccccggctat	tgccagggag	gtggctgggt	3660
tcaaggaccg	gcgacttgct	aaggaccccg	gaagggggag	gggggtgggg	tgccctccacg	3720
tgccagcggg	gacttggggg	agtccttggg	gatggcaaaa	acctgacctg	tgaaggggac	3780
acagtttggg	ggttgagggg	aagaaggttt	gggggttctg	ctgtgccagt	ggagagggaag	3840
ctgataagct	gataacctgg	gcgctggagc	caccacttat	ctgccagagg	ggaagcctct	3900
gtcacaccag	gattgaagtt	tggccggaga	agtggatgct	ggtagctggg	ggtgggggtg	3960
gcacacggca	gcaggattga	atgaaggcca	gggagggcag	acctgagtgc	ttgcatgggt	4020
ggggacagga	aggacgagct	ggggcagaga	cgtgggggatg	aaggaaagctg	tccttccaca	4080
gccacccctc	tccctccccg	cctgactctc	agcctggcta	tctgttctag	aatgtcctgc	4140
ctggctgtgg	cttctcctgt	ccctgctgtc	gctccctctg	ggcctcccag	tcctggggcg	4200
cccacacgc	ctcatctgtg	acagccgagt	ctcgagagg	tacctcttgg	aggccaagga	4260
ggccgagaat	atcacggtga	gacccttcc	ccagcacatt	ccacagaact	cacgctcagg	4320
gcttcaggga	actcctccca	gatccaggaa	cctggcactt	ggtttggggg	ggagttggga	4380
agctagacac	tgccccccca	cataagaata	agtcctgggtg	ccccaaacca	tacctggaaa	4440
ctaggcaagg	agcaaaagcca	cgagatccca	cggcctgtgg	gccagggcca	gagccttcag	4500
ggacccttga	ctccccgggc	tgtgtgcatt	tcagacgggc	tgtgctgaac	actgcagctt	4560

-98-

gaatgagaat	atcactgtcc	cagacaccaa	agttaatttc	tatgcctgga	agaggatgga	4620
ggtgagttcc	tttttttttt	tttttccttt	cttttggaga	atctcatttg	cgagcctgat	4680
tttggatgaa	agggagaaatg	atcgagggaa	aggtaaaatg	gagcagcaga	gatgaggtcg	4740
cctgggagca	gaggctcacg	tctataatcc	caggctgaga	tggccgagat	gggagaattg	4800
cttgagccct	ggagtttcag	accaacctag	gcagcatagt	gagatcccc	atctctacaa	4860
acatttaaaa	aaattagtca	ggtgaagtgg	tgcatggtgg	tagtcccaga	tatttggag	4920
gctgagggcg	gaggatcgct	tgagcccagg	aatttgaggc	tgcatgagc	tgtgatcaca	4980
ccactgcact	ccagcctcag	tgacagagt	aggccctgtc	tcaaaaaaga	aaagaaaaaa	5040
gaaaaataat	gagggctgta	tggaatacat	tcattattca	ttcactcact	cactcactca	5100
ttcatttcatt	cattcattca	acaagtctta	ttgcatacct	tctgtttgct	cagcttggtg	5160
cttggggctg	ctgaggggca	ggagggagag	ggtgacatgg	gtcagctgac	ttccagagtc	5220
cactccctgt	aggctcgggca	gcaggccgta	gaagtctggc	agggcctggc	cctgctgtcg	5280
gaagctgtcc	tgcggggcca	ggccctgttg	gtcaactott	cccagccgtg	ggagccccctg	5340
cagctgcacg	ctgataaagc	cgtagtggtc	cttcgcagcc	tcaccactct	gcttcgggct	5400
ctgggagccc	aggtgagtag	gagcggacac	ttctgcttgc	cctttctgta	agaaggggag	5460
aagggctctg	ctaaggagta	caggaactgt	ccgtattcct	ttcccttctg	tggcactgca	5520
gcgacctcct	gttttctcct	tggcagaagg	aagccatctc	ccctccagat	gcggcctcag	5580
ctgtccact	ccgaacaatc	actgctgaca	ctttccgcaa	actcttccga	gtctactcca	5640
atttccctcg	gggaaagctg	aagctgtaca	caggggaggc	ctgcaggaca	ggggacagat	5700
gacgtacaag	taagaattca	ctcctcaggt	gcaggctgcc	tatcagaagg	tgggtggctgg	5760
tgtggccaat	gccctggctc	acaaatacca	ctgagatctt	tttccctctg	ccaaaaatta	5820
tggggacatc	atgaagcccc	ttgagcatct	gacttctggc	taataaagga	aattttatatt	5880
cattgcaata	gtgtgttgga	attttttgtg	tctctcactc	ggaaggacat	atgggagggc	5940
aaatcatttta	aaacatcaga	atgagtattt	ggttttagagt	ttggcaacat	atgccatatg	6000
ctggctgcca	tgaacaaagg	tggctataaa	gaggtcatca	gtatatgaaa	cagccccctg	6060
ctgtccattc	cttattccat	agaaaagcct	tgacttgagg	ttagattttt	tttatatttt	6120
gttttgtgtt	atttttttct	ttaacatccc	taaaattttc	cttacatgtt	ttactagcca	6180
gatttttctc	cctctcctga	ctactcccag	tcatagtgt	ccctcttctc	ttatgaagat	6240
ccctcgacct	gcagcccaag	cttgcatgcc	tgcaagtcga	ctctagtggg	ttcccccgcc	6300
cgtatcccc	agggtgtctg	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	6360
cccgctgccac	cttcccctg	cccggtctgt	ccccgcacgc	tgccggctcg	gggatcgggg	6420
gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	gcgggggagg	6480
gacgtaatta	catccctggg	ggcttttggg	gggggctgtc	cccgtagcgc	gatccgcggc	6540
cccgtatccc	ccaggtgtct	gcaggtcaa	agagcagcga	gaagcgttca	gaggaaagcg	6600
atcccgctgc	accctccccg	tgcccggtct	gtccccgcac	gctgcccgtc	cggggatgctg	6660
gggggagcgc	cggaccggag	cggagccccg	ggcggtctgc	tgctgcccc	tagcggggga	6720
gggacgtaat	tacatccctg	ggggcttttg	gggggggctg	ttcccgtgag	cggatccgcg	6780
gccccgtatc	cccaggtgtg	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaag	6840
cgatcccggtg	ccaccttccc	cgtgcccggg	ctgtccccgc	acgctgcccg	ctcggggatg	6900
cggggggagc	gcccggaccg	agcggagccc	cggggcggtc	gctgctgccc	cctagcgggg	6960
gagggacgta	attacatccc	tgggggcttt	tggggggggg	tgtccccgtg	agcggatccg	7020
cggccccgta	tccccaggt	gtctgcaggc	tcaaaagcga	gcgagaagcg	ttcagaggaa	7080
agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	7140
tgccggggga	gcgcgggacc	ggagcggagc	cccggggcgc	tcgctgctgc	cccctagcgg	7200
gggagggacg	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	7260
cgcggggccc	tatccccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	7320
aaagcgatcc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	7380
gatgccccgg	gagcgccgga	ccggagcggg	gccccggggc	gctcgctgct	gccccctagc	7440
gggggagggg	cgtaatatac	tcccgggggg	ctttgggggg	gggctgtccc	cgtgagcggg	7500
tccgcggccc	cgtatcccc	aggtgtctgc	aggtcaaag	agcagcgaga	agcgttcaga	7560
ggaaagcgat	cccgtgccac	cttcccctgt	cccgggctgt	ccccgcacgc	tgccggctcg	7620
gggatgcccc	gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	7680
gcgggggagg	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtagcgc	7740
gatccgcggg	gctgcaggaa	ttcgtaatca	tggctcatagc	tggttccctg	gtgaaattgt	7800
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctggggg	7860
gcctaattgag	tgagctaact	cacattaatt	cggttgcgct	cactgcccgc	tttccagtcg	7920
ggaaacctgt	cgtgccagct	gcattaatga	atcgcccaac	gcgcggggag	aggcggtttg	7980
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgctgctcgg	cgttcggctg	8040
cggcgagcgg	tatcagctca	ctcaaaaggc	gtaatacggg	tatccacaga	atcaggggat	8100
aacgcaggaa	agaaatgtg	agcaaaaggc	cagcaaaagg	ccaggaaaccg	taaaaaggcc	8160
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	8220
tcaagtcaga	tgtggcgaaa	cccagacagg	ctataaagat	accagggcgt	tccccctgga	8280
agctccctcg	ggcgctctcc	tgctccgacc	ctgcgcttta	ccgataacct	gtccgccttt	8340
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	8400
taggtcgttc	gctccaagct	gggctgtgtg	acccaacccc	ccgttcagcc	cgaccgctgc	8460
gccttatccg	gtaactatcg	tcttgagtc	aaccggtaaa	gacacgactt	atcgccactg	8520
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	8580

-99-

```

ttgaagtgggt ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg 8640
ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaacacc 8700
gctggtagcg gtggtttttt tggttgcaag cagcagatta cgcgcagaaa aaaaggatct 8760
caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaaacga aaactcacgt 8820
taagggatct tgggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa 8880
aaatgaagtt ttaaataaat ctaaagtata tatgagtaaa cttgggtctga cagttacca 8940
tgcttaatac gtgaggcacc tatctcagcg atctgtctat ttcgttcac catagtgtgc 9000
tgactccccg tctgttagat aactacgata cgggaggggt taccatctgg cccagtgct 9060
gcaatgatac cgcgagaccc acgtcacccg gctccagatt tatcagcaat aaaccagcca 9120
ccgggaaggg ccgagcgagc agtggtcctt gcaactttat ccgcctccat ccagtctatt 9180
aattgttgcc ggggaagctag agtaagtagt tccgagttta atagtgtgag caacgttgtt 9240
gccattgcta caggcatcgt ggtgtcacgc tctgtctgtg gtatggcttc attcagctcc 9300
tgctcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc 9360
tccttcgggtc tccgagcgt agtcagaagt aagttggccg cagtgttacc actcatgggt 9420
atggcagcac tgcataattc tcttactgtc atgcatccg taagatgctt ttctgtgact 9480
gggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc 9540
ccggcgctcaa tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt 9600
ggaaaacggt cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg 9660
atgtaaccca ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct 9720
gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa 9780
tggtgaatac tcatactctt cctttttcaa tatattgaa gcatttatca ggttattgt 9840
ctcatgagcg gatacatatt tgaatgtatt tagaaaaata acaaaatagg ggttccgcgc 9900
acatttcccc gaaaagtgcc acctgacgta gttaaacaaa aaaagccgc cgaagcgggc 9960
tttattacca agcgaagcgc cattcgccat tcaggctcgc caactgttgg gaagggcgat 10020
cgggtcggggc ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct gcaagggcat 10080
taagttgggt aacgccaggg ttttcccagt cacgacgttg taaaacgacg gccagtcctg 10140
aatacgactc acttaaggcc ttgactagag ggtcgacggt atacagacat gataagatac 10200
attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa 10260
atttgtgatg ctattgcttt atttgaatac attataagct gcaataaaca agttgggggtg 10320
ggcgaagaac tccagcatga gatccccgcg ctggaggatc atccagccgg cgtcccgaa 10380
aacgattccg aagcccaacc tttcatagaa ggcgcgggtg gaatcgaaat ctcgtagcac 10440
gtgtcagtc tgcctctcgg ccacgaagtg cacg 10474

```

<210> 126

<211> 6119

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeoeGFP Plasmid

<400> 126

```

cagttgcccgg ccgggtcgcg caggcggaac tcccccccc acggctgctc gccgatctcg 60
gtcatggccg gcccgaggc gtcccgaag ttctgtggaca cgacctcga ccactcggcg 120
tcagctcgt ccaggcccg caccacacc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgtgatgaa cagggtcacg tctgtccgga ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagccgg tccgtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaagca ggcttcaatc ctgcagagaa gcttgggctg caggctcagg 420
gatcttcata agagaagagg gacagctatg actgggagta gtcaggagag gaggaaaaat 480
ctggctagta aaacatgtaa ggaaaatttt agggatgtta aagaaaaaaa taacacaaaa 540
caaaatataa aaaaaatcta acctcaagtc aaggcttttc tatggaataa ggaatggaca 600
gcagggggct gtttcatata ctgatgacct ctttatagcc acctttgttc atggcagcca 660
gcatattggc tatgttgcca aactctaaac caaatactca ttctgatgtt ttaaattgatt 720
tgccctccca tatgtccttc cgagttagag acacaaaaaa ttccaacaca ctattgcaat 780
gaaaataaat ttcctttatt agccagaagt cagatgctca aggggcttca tgatgtcccc 840
ataatttttg gcagagggaa aaagatctca gtgggtattt tgagccaggg cattggccac 900
accagccacc acctcttgat aggcagcctg cacctgagga gtgaattctt acttgtacag 960
ctcgtccatg ccgagagtga tcccggcggc ggtoacgaac tccagcagga ccatgtgatc 1020
gcgcttctcg ttgggtctt tgcacaggc ggactgggtg ctcaggtagt ggttgtcggg 1080
cagcagcacg gggcgtcgc cgtatggggg gttctgctgg tagtggtcgg cgagctgcac 1140
gctgccgtcc tcatgttgt ggcggatctt gaagttcacc ttgatgccgt tcttctgctt 1200
gtcggccatg atatagcgt tgtggctgtt gtagttgtac tccagcttgt gccccaggat 1260
gttgccgtcc tccttgaagt cgatgccctt cagctcgatg cggttcacca gggtgtcgcc 1320
ctcgaacttc acctcgcgcg gggctctgta gttgccgtcg tccttgaaga agatgggtgc 1380
ctcgtggacg tagcctcgg gcacggcgga cttgaagaag tctgtgctgt tcatgtggtc 1440
ggggtagcgg ctgaagcact gcacgccgta ggtcagggtg gtcacgaggg tgggcccagg 1500

```

-100-

cacggggcagc	ttgcccgggtg	tgcagatgaa	cttcagggtc	agcttgccgt	aggtggcatc	1560
gccctcgccc	tgcgcggaca	cgctgaactt	gtggccggtt	acgtcgccgt	ccagctcgac	1620
caggatgggc	accaccccg	tgaacagctc	ctcgcccttg	ctcaccatgg	tggcgaattc	1680
tttgccaaaa	tgatgagaca	gcacaacaac	cagcacgttg	cccaggagct	gtaggaaaaa	1740
gaagaaggca	tgaacatggt	tagcagaggc	tctagagccg	ccggtcacac	gccagaagcc	1800
gaaccccgcc	ctgccccgtc	ccccccgaag	gcagccgtcc	ccctgcggca	gccccgaggc	1860
tgagatgga	gaaggggacg	gcgccgaggc	gacgcacgaa	ggccctcccc	gccccatttc	1920
ttcctgccc	cgccgcaccg	cttgcgccgc	gcccgttaga	gggggtgcgg	cgccgcctcc	1980
cagatttcgg	ctccgcagca	tttgggacaa	aggaagtccc	tgcgccctct	cgcacgatta	2040
ccataaaagg	caatggctgc	ggctcgccgc	gcctcgacag	ccgcccggcg	tccggggccg	2100
ccgcgcccct	cccccgagcc	ctccccggcc	cgaggcggcc	ccgccccgcc	cgccaccccc	2160
acctgcccgc	accccccgcc	cggcacggcg	agcccccgcg	cacgcccccg	acggagcccc	2220
gcacccgaag	ccgggcccgt	ctcagcaact	cggggagggg	gggtgcaggg	gggggttacag	2280
cccgaccgcc	gcgcccacac	ccccctgctc	cccccccacg	cacacacccc	gcacgcagcc	2340
tttggtcccc	tgcgagcccc	cccgccaccg	ggggcaccgc	ccccggccgc	gctccccctc	2400
cgcacacgcg	gagcgcacaa	agcccccgcg	cgcgcccgca	gcgctcacag	cccgccgggca	2460
gcgcggggcg	cacgcggcgc	tccccacgca	cacacacacg	cacgcacccc	ccgagccgct	2520
cccccccgca	caaagggccc	tccccgagcc	ctttaaggct	ttcacgcagc	cacagaaaag	2580
aaacgagccg	tcattaaacc	aagcgctaatt	tacagcccgg	aggagaaggg	ccgtcccgcg	2640
cgctcacctg	tgggagtaac	gcggtcagtc	agagccgggg	cgggcgggcg	gaggccggcg	2700
ggagcggggc	acggggcgaa	ggcaacgcag	cgactcccg	ccgcccgcgc	cttcgctttt	2760
tatagggccg	ccgcccggcg	cgcctcgcca	taaaaggaaa	ctttcggagc	gcgcccgtct	2820
gattggctgc	cgccgcaccc	ctccgcctcg	ccccgcccc	ccctcgccc	cgcccccgcc	2880
cgctggcgcg	cgcccccccc	cccccccgcc	ccccctgctg	cacaaaaata	ttaaaaataa	2940
aataaataca	aaattggggg	tggggagggg	ggggagatgg	ggagagtga	gcagaacgtg	3000
gggctcacct	cgacccatgg	taatagcgat	gactaatacg	tagatgtact	gccaaagtagg	3060
aaagtcccat	aaggtcatgt	actgggcata	atgccaggcg	ggccatttac	cgctattgac	3120
gtcaataggg	ggcgtaactg	gcataatgata	cacttgatgt	actgccaagt	gggcagttta	3180
ccgtaaatag	tccacccatt	gacgtcaatg	gaaagtccct	attggcgcta	ctatgggaac	3240
atacgtcatt	attgacgtca	atgggcgggg	ctcgttgggc	ggtcagccag	gcggggcatt	3300
taccgtaagt	tatgtaacgc	ggaaactccat	atatgggcta	tgaactaatg	accccgtaat	3360
tgattactat	taataactag	aggatccccg	ggtaccgagc	tcgaattcgt	aatcatgggc	3420
atagctgttt	cctgtgtgaa	attgttatcc	gtcacacatt	ccacacaaca	tacgagccgg	3480
aagcataaag	tgtaaaagcct	gggtgacctc	atgagtgagc	taactcacat	taattgcggt	3540
gcgctcactg	cccgccttcc	agtcgggaaa	cctgtcgtgc	cagctgcatt	aatgaatcgg	3600
ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctct	tccgcttcc	cgctcactga	3660
ctcgctgcgc	ctcggtcgtc	ggctgcggcg	agcggtatca	gctcactcaa	aggcggtaat	3720
acgggtatcc	acagaatcag	gggataacgc	aggaaaagaac	atgtgagcaa	aaggccagca	3780
aaaggccagg	aaccgtaaaa	aggccgcgct	gctggcgctt	ttccataggc	tccgcccccc	3840
tgacgagcat	cacaaaaatc	gacgtcaag	tcagagggtg	cgaaccgga	caggactata	3900
aagataaccag	cggtttcccc	ctggaagctc	cctcgtgcgc	tctcctgttc	cgacctgcc	3960
gcttaccgga	tacctgtccg	cctttctccc	ttcggaagc	gtggcgcttt	ctcatagctc	4020
acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	aagctgggct	gtgtgcacga	4080
acccccgctt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaacc	4140
ggtaagacac	gacttatcgc	cactggcagc	agccactgg	aacaggatta	gcagagcgag	4200
gtatgtaggc	gggtgtacag	agttcctgaa	gtggtggcct	aactacggct	acactagaag	4260
gacagtattt	gggtatctgc	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4320
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgggtgg	ttttttgttt	gcaagcagca	4380
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atcttttcta	cggggtctga	4440
cgctcagtg	aacgaaaact	cacgttaagg	gattttgggt	atgagattat	caaaaaggat	4500
cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	tcaatctaaa	gtatatatga	4560
gtaaaacttg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4620
tctatttcgt	tcatccatag	ttgctgact	ccccgtcgtg	tagataacta	cgatacggga	4680
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	gaccacgct	caccggctcc	4740
agattttatca	gcaataaacc	agccagccgg	aagggccgag	cgcagaagtg	gtcctgcaac	4800
tttatccgcc	tccatccagt	ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	4860
agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	atcgtgggtg	cacgctcgtc	4920
gtttgggtatg	gcttcattca	gctccggttc	ccaacgatca	aggcgagtta	catgatcccc	4980
catgttgtgc	aaaaaagcgg	ttagctcctt	cggtcctccg	atcgttgtca	gaagtaagtt	5040
ggccgcagtg	ttatcactca	tgggttatggc	agcactgcac	aattctctta	ctgtcatgcc	5100
atccgtaaga	tgtctttctg	tgactggtga	gtactcaacc	aagtcattct	gagaatagtg	5160
tatgcggcga	ccgagttgct	cttgcccggc	gataatacgg	gataatacgg	gcacacatag	5220
cagaacttta	aaagtgtcga	tcattggaaa	acgttcttcg	gggcgaaaac	tctcaaggat	5280
cttaccgctg	ttgagatcca	gttcgatgta	acccactcgt	gcacccaact	gatcttcagc	5340
atcttttact	ttctcgggtg	ttctcgggtg	agcaaaaaaca	ggaaggcaaa	atgccgcaaa	5400
aaagggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttccttt	ttcaatatta	5460
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatac	atatttgaat	gtatttagaa	5520

-101-

```

aaataaaciaa ataggggttc cgcgacatt tccccgaaaa gtgccacctg acgtagttaa 5580
caaaaaaaag cccgccgaag cgggctttat taccaagcga agcgccattc gccattcagg 5640
ctgcccgaact gttgggaagg gcgatcgggtg cgggcctctt cgctattacg ccagctggcg 5700
aaaggggggat gtgctgcaag gcgattaagt tgggtaacgc cagggttttc ccagtcacga 5760
cggttgtaaaa cgacggccag tccgtaatac gactcaacta aggccttgac tagagggtcg 5820
acgggtataca gacatgataa gatacattga tgagtttgga caaaccacaa ctagaatgca 5880
gtgaaaaaaa tgctttatct gtgaaatttg tgatgctatt gctttatttg taaccattat 5940
aagctgcaat aaacaagtgt ggggtgggcga agaactccag catgagatcc ccgcgctgga 6000
ggatcatcca gccggcgctc cggaaaacga ttccgaagcc caacctttca tagaaggcgg 6060
cggtggaatc gaaatctcgt agcacgtgtc agtctgtctc ctccggccagc aagtgcacg 6119

```

<210> 127

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXLamInt Plasmid (Wildtype Integrase)

<400> 127

```

gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccataatat ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 120
ccaacgaccc cgcgccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggacttttcca ttgacgtcaa tgggtggact atttacggta aactgccacac ttggcagtag 240
atcaagtgtg tcatatgcc agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtag atgaccttat gggacttttc tacttggcag tacatctacg 360
tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctcccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
gcgagtggtgg cggggggggg ggggggcgcgc gccaggcggg gcggggcggg gcgagggggc 540
gggcgggggg aggcggagag gtgcccgcgc agccaatcag agcggcgcgc tccgaaagt 600
tccttttatg gcgagcgggc ggccgagcgc gccctataaa aagcgaagcg cgcggcgggc 660
gggagtcgct gcgttgccct cgcctccgtg cccgctccgc gccgcctcgc gccgcccgc 720
ccggctctga ctgaccgcgt tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780
gggctgtaat tagcgcttg ttaaatgac gctcgtttct tttctgtggc tgcgtgaaag 840
ccttaaaggg ctccgggagg gccctttgtg cgggggggag cggctcgggg ggtgcgtgag 900
tgtgtgtgtg cgtggggagc gcccgctgag gcccgcgctg cccggcgggt gtgagcgtg 960
cgggcccggc gcggggcttt gtgcgctccg cgtgtgcgcg aggggagcgc ggcggggggc 1020
gggtcccgcg ggtgcggggg ggctgcgagg ggaacaaagg ctgcgtgcgg ggtgtgtgag 1080
tggggggggtg agcagggggg gtgggcgcgg cggtcgggct gtaaccccc cctgcacccc 1140
cctccccgag ttgctgagca cggcccggct tcgggtgcgg ggctccgtgc ggggctgtgc 1200
gcggggctcg ccgtgccggg cgggggggtg cggcaggtgg ggggtgccgg cggggcgggg 1260
ccgctcggg ccggggaggg ctccgggggag gggcgcgggc gccccggagc gccggcgggc 1320
gtcgagggcg ggcgagccgc agccattgcc ttttatggta atcgtgcgag agggcgagc 1380
gacttctctt gtcctcaatc tggcggagcc gaaatctggg aggcgcggcc gcaccccctc 1440
tagcggcgcg ggcgaaagc gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt 1500
cgtgcgtcgc cgcgcgcgcg tccccttctc catctccagc ctccgggctg ccgcaggggg 1560
acggctgcct tcggggggga cggggcaggg cgggggtcgg cttctggcgt gtgaccggcg 1620
gctctagagc ctctgctaac catgttcag ccttcttctt tttctacag ctccgtggca 1680
acgtgctggt tgtgtgtgtg tctcatcatt ttggcaaaga attcatggga agaaggcgaa 1740
gtcatgagcg ccgggattta ccccctaacc tttatataag aaacaatgga tattactgct 1800
acagggaccc aaggacgggt aaagagtttg gattaggcag agacaggcga atcgcaatca 1860
ctgaagctat acaggccaac attgagttat ttccaggaca ccaacacaag cctctgacag 1920
cgagaatcaa cagtataat tccgttacgt tacattcatg gcttgatcgc tacgaaaaaa 1980
tcctggccag cagaggaatc aagcagaaga cactcataaa ttacatgagc aaaattaaag 2040
caataaggag ggtctgcct gatgctccac ttgaagacat caccacaaaa gaaattgcgg 2100
caatgctcaa tggatacata gacgagggca aggcggcgct agccaagtta atcagatcaa 2160
cactgagcga tgcattccga gaggcaatag ctgaaggcca tataacaaca aacctgtcg 2220
ctgccactcg cgcagcaaaa tcagaggtaa ggagatcaag acttacggct gacgaatacc 2280
tgaaaaattt tcaatcatcac gaatcatcac catgttggct cagacttgca atggaactgg 2340
ctgttggttac cgggcaacga gttggtgatt tatgcgaaat gaagtggctc gatatcgtag 2400
attgatatct ttatgtcgag caaagcaaaa caggcgtaaa aattgccaat ccaacagcat 2460
tgcatctcga tgctctcgga atatcaatga aggaacact tgataaatgc aaagagattc 2520
ttggcgggaga aaccataatt gcatctactc gtgcgaacc gctttcatcc ggcacagtat 2580
caaggtatct tatgcgcgca cgaaaagcat caggtctttc cttcgaaggg gatccgccta 2640
cctttcacga tttgtcgagt ttgtctgcaa gactctatga gaagcagata agcgataagt 2700
ttgctcaaca tcttctcggt cataagtcgg acaccatggc atcacagtat cgtgatgaca 2760
gaggcagggg gtgggacaaa attgaaatca aataagaatt cactcctcag gtgcaggctg 2820

```

-102-

```

cctatcagaa ggtggtggct ggtgtggcca atgccctggc tcacaaatac cactgagatc 2880
tttttccctc tgccaaaaat tatggggaca tcatgaagcc ccttgagcat ctgacttctg 2940
gctaataaag gaaattttatt ttcattgcaa tagtgtgttg gaattttttg tgtctctcac 3000
tcggaaggac atatgggagg gcaaatcatt taaaacatca gaatgagtat ttggtttaga 3060
gtttggcaac atatgccata tgctggctgc catgaacaaa ggtggctata aagaggtcat 3120
cagtatatga aacagccccc tgctgtccat tccttattcc atagaaaagc cttgacttga 3180
ggtttagattt tttttatatt ttgttttgtg ttattttttt ctttaacatc cctaaaattt 3240
tccttacatg ttttactagc cagatttttc ctctctctct gactactccc agtcatagct 3300
gtccctcttc tcttatgaag atccctcgac ctgcagccca agcttggcgt aatcatggtc 3360
atagctgttt cctgtgtgaa attgttatcc gtccacaatt ccacacaaca tacgagccgg 3420
aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat taattgcgtt 3480
gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagcggatcc gcatctcaat 3540
tagtcagcaa gcccctaac cgccctaac cgccctaac tccgccagt 3600
tccgcccatt ctccgcccga ggctgacta tttatgcaga ggccgaggcc 3660
gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctaggcttt 3720
tgcaaaaagc taacttgttt attgcagctt ataattggtt caaataaagc aatagcatca 3780
caaatttcac cctgttgtaa tttttttcac tgcttctag ttgtggtttg tccaaactca 3840
tcaatgtatc ttatcatgtc tggatccgct gcattaatga atcggccaac gcgcggggag 3900
agggcggttg cgtattgggc gctcttccgc ttctcgtctc actgactcgc tgcgctcggg 3960
cggtcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg gtaatacggg 4020
atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg 4080
taaaaaggcc gcgttgctgg cggtttttcca taggtccgcg cccctgacg cccctgacg 4140
aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt 4200
tccccctgga agctccctcg tgctccgacc ctgcccgtta ccgggacctt 4260
gtccgccttt ctcccttcgg gaagcgtggc gctccaagct gcttgcgacc tgctcacgct gtaggtatct 4320
cagttcgggt taggtcggtt gtaactatcg ggtcgtgtg caggaacccc ccgttcagcc 4380
cgaccgctgc gccttatcgg gtaactatcg ggtcgtgtg aacctcggtt gacacgactt 4440
atcgccactg gcagcagcca gtgtaacag ggcctaacta gattagcaga gcgaggtatg taggcgggtg 4500
tacagagttc ttgaagtggg ttaccttcgg cggctacact agaaggacag tatttgggtat 4560
ctgcgctctg ctgaagccag gtggtttttt aaaaagagtt ggtagctctt gatccggcaa 4620
acaaaccacc cgtggtagcg tggttgcaag cagcagatta cgcgagaaa 4680
aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga 4740
aaactcacgt taagggattt tggatcatgag attatcaaaa aggatcttca cctagatcct 4800
tttaaattaa tgaatgaagt gtgaggcacc cttaaagtata tatgagtaaa cttggtctga 4860
cagttaccaaa tgcttaatca gtgagtcacc tatctcagcg atctgtctat ttcggttcac 4920
catagttgcc tgactccccg tcgtgtagat aactacgata cgggaggggt taccatctgg 4980
ccccagtgct gcaatgatac acgctcaccg gctccagatt tatcagcaat 5040
aaaccagcca gccggaaggg ggaagctag gcaactttat ccgcctccat 5100
ccagtcctatt aattgttgcc aggtgtcacg tcgccagtta atagtttgcg 5160
caacgttggt gccattgcta agttacatga tcgctggttg gtatggcttc 5220
attcagctcc ggttcccaac gatcaaggcg tccccatgt tgtgcaaaaa 5280
agcgggttagc tccttcgggtc tgtcagaagt aagttggcgg cagtgttatc 5340
actcatgggt atggcagcac tcttatgtgc atgccatccg taagatgctt 5400
ttctgtgact ggtgagtagt attctgagaa tagtgtatgc ggcgaccgag 5460
ttgctcttgc cggcgctcaa taccgcgcca catagcagaa ctttaaaagt 5520
gctcatcatt ggaaaacgtt ctccggggcg aggatcttac cgctggtgag 5580
atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt ttactttcac 5640
cagcgtttct ggggtgagcaa aaacaggaag gcaaaaaagg gaataagggc 5700
gacacggaaa tggtgaatac tcatactctt cctttttcaa tattattgaa gcatttatca 5760
gggttattgt ctcagagcgc gatacatatt tgaatgtatt tagaaaaata aacaaatagg 5820
ggttccgcgc acatttccc gaaaagtggc acctg 5855

```

<210> 128

<211> 303

<212> DNA

<213> Artificial Sequence

<220>

<223> Human FER-1 Promoter

<400> 128

```

tccatgacaa agcacttttt gagcccaagc ccagcctagc tcgagctaaa cgggcacaga 60
gacgccaccg ctgtcccaga ggcagtcggc taccggtccc cgctcccag ctcccgcaga 120
gcgcgcgagg gcctccagcg gccgcccctc ccccacagca ggggcggggg cccgcgcccc 180
ccggaaggag cgggctcggg gcgggcgggc ctgattggcc ggggcggggc tgacgccgac 240
gcggctataa gagaccacaa gcgacccgca gggccagacg ttcttcgccc agagtcgggt 300
acc 303

```

-103-

<210> 129
 <211> 6521
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pIRES-BSR Plasmid

<400> 129
 tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattggcta 60
 ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120
 aatatgaccg ccattgttggc attgattatt gactagttat taatagtaat caattacggg 180
 gtccattagt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
 gcctggctga ccgcccacg acccccgccc attgacgtca ataatgacgt atgttcccat 300
 agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaaactgc 360
 ccacttggga gtacatcaag tgtatcatat gccaaagtccg cccctattg acgtcaatga 420
 cggtaaatgg ccgcctggc attatgcccc gtacatgacc ttacgggact ttccacttg 480
 gcagtacatc tacgtattag tcatcgctat taccatgggt atgcggtttt ggcagtacac 540
 caatgggctg ggtatgcgtt ttgactcacg gggatttcca agtctccacc ccattgacgt 600
 caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactg 660
 cgatcgcccc cccggttgac gcaaattggc ggtaggcgtg tacgggtggga ggtctatata 720
 agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgagg agtttatcac 780
 agttaaatgg ctaacgcagt cagtgtcttc gacacaacag tctcgaactt aagctgcagt 840
 gactctctta aggtagcctt gcagaagtgt gtcgtgaggc actggggcagg taagtatcaa 900
 gggtacaaga cagggtttaag gagaccaata gagactgctt tctcgaactt aagctgcagt 960
 ctgctctctt ttagtaggac ctattgggtc ttaaggctag agggcggcca attccgcccc 1020
 aggtgtccac tcccagttca attacagctc gcacatcctt cactttgcct ttctctccac 1080
 ataggctagc ctcgagaatt cacgcgtcga cgttactggc tactgacatc cactttgcct 1140
 tctccctccc cccccctaa cccatatttg cgggtctctt ttaaggctag agggcggcca 1200
 tttgtctata tgtgattttc cactatattg aggggtcttt cgggtctctg gcaatgtgag 1260
 cctggccctg tctcttgac gaaggaagca caggcagcgg aacccccac ctggcgacag 1320
 caaggtctgt tgaatgtcgt caggtgtgata agagtcaaa tggctctcct caagcgtatt 1380
 acgtctgtag cgaccctttg agatcacctt aagagtcaaa tggctctcct caagcgtatt 1440
 ggcaaaaagc tagttgtgga ctgaaggatg cccagaaggt accccattgt atgggactct 1500
 gtgagttgga tagttgtgga ctgaaggatg cccagaaggt accccattgt atgggactct 1560
 ctgaaggatg cccagaaggt accccattgt atgggactct atctggggcc 1620
 tggctttacat gtgttttagt acgatgataa gaattagtag gtgggagcgg caattcgtac 1680
 gtgggttttc tttgaaaaac acgatgataa gaattagtag gtgggagcgg caattcgtac 1740
 acatttctca acaagatcta atagtagtag gtgggagcgg caattcgtac 1800
 atgaggataa taaacatcat cggcagtaga tatataggac gagtaactgt ttgtgcagaa 1860
 cggcagtaga tatataggac gagtaactgt ttgtgcagaa gattttagct 1920
 ttggtagtgc agtttgcgaat agagttagat ccagattggt ccaaatatac 1980
 ctatttctga cgaagtagat ccagattggt ccaaatatac ccaaatatac 2040
 agttgatttc agactatgca ctcatccac ccgcttccct ttagtgaggg 2100
 tcaaaactac gattgaagaa ccgcttccct ccaaatatac ccaaatatac 2160
 ataccaagct tggcggggcg attgatgagt ttggacaaac ccaaatatac 2220
 gataagatac attgatgagt ttggacaaac ccaaatatac ccaaatatac 2280
 tatttgtgaa atttgtgatg agtttgcgaat aacaattgca tttttaaagc 2340
 agtttgcgaat aacaattgca tttttaaagc aagtaaaacc tggcgtaata 2400
 tttttaaagc aagtaaaacc tggcgtaata ggcgaatgga cgcgccctgt 2460
 ggcgaatgga cgcgccctgt gctgtgacgc tacacttgcc gttcgccggc 2520
 ttctcgccac gttcgccggc cactctgacc tagacgggtt caaactggaa 2580
 tccgatttag agctttacgg atcgccctga actcctgttc cggatttcgg 2640
 gtagtgggac actcctgttc cggatttcgg aacaaaatat cggtatttca 2700
 ttgatattata agggattttg cggatttcgg aacaaaatat cggtatttca 2760
 aaatatttaa cgcgaatttt cggatttcgg aacaaaatat cggtatttca 2820
 tttctcctta cgcctctgtg agaggaactt ttaggggtgt aattagtcag 2880
 agctgtggaa ttaggggtgt aattagtcag agcatgcac ctaactccgc 2940
 gtatgcaaag catgcacatc agcatgcac ctaactccgc gcagaggccg 3000
 cagcaggcag aagtagcaa ctaactccgc gcagaggccg ggaggccctg 3060
 taactccgc ctaactccgc gcagaggccg ggaggccctg gcttttgcaa 3120
 gactaatttt ttttatttat gcttttgcaa aaagcttgat tcttctgaca 3180
 agtagtgagg aggccttttt gcttttgcaa aaagcttgat tcttctgaca 3240
 agtagtgagg aggccttttt gcttttgcaa aaagcttgat tcttctgaca 3300
 agtagtgagg aggccttttt gcttttgcaa aaagcttgat tcttctgaca 3360
 agtagtgagg aggccttttt gcttttgcaa aaagcttgat tcttctgaca 3420
 agtagtgagg aggccttttt gcttttgcaa aaagcttgat tcttctgaca 3480

-104-

caacagtctc	gaacttaagg	ctagagccac	catgattgaa	caagatggat	tgcacgcagg	3540
ttctccggcc	gcttgggtgg	agaggctatt	cggctatgac	tgggcacaac	agacaatcgg	3600
ctgctctgat	gccgccgtgt	tccggctgtc	agcgcagggg	cgcccggttc	tttttgtcaa	3660
gaccgacctg	tccggtgccc	tgaatgaact	gcaggacgag	gcagcgcgcc	tatcgtggct	3720
ggccacgacg	ggcgttcctt	gcgcagctgt	gctcgacggt	gtcactgaag	cggaagggga	3780
ctgggtgcta	ttggggcgaag	tgccggggca	ggatctcctg	tcactctacc	ttgctcctgc	3840
cgagaaaagta	tccatcatgg	ctgatgcaat	gcggcggtcg	catacgcttg	atccggtac	3900
ctgcccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	ggatggaagc	3960
cggctctgtc	gatcaggatg	atctggacga	agagcatcag	gggctcgccg	cagccgaact	4020
gttcgccagg	ctcaaggcgc	gcatgcccg	cggcgaggat	ctcgtcgtga	cccatggcga	4080
tgcctgcttg	cgaatatca	tgggtgaaaa	tggccgcttt	tctggattca	tcgactgtgg	4140
ccggctgggt	gtggcggacc	gctatcagga	catagcgctg	gctacccgtg	atattgctga	4200
agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggtatcg	ccgctcccga	4260
ttcgcagcgc	atcgcttct	atcgcttct	tgacgagttc	ttctgagcgg	gactctgggg	4320
ttcgaaatga	ccgaccaagc	gacgcccac	ctgccatcac	gatggccgca	ataaaatc	4380
tttattttca	ttacatctgt	gtgttggttt	tttgtgtgaa	tcgatagcga	taaggatccg	4440
cgtatggtgc	actctcagta	caatctgctc	tgatgccgca	tagttaagcc	agccccgaca	4500
ccgcgcaaca	cccgtcgacg	cgccctgacg	ggcttgtctg	ctcccgccat	ccgcttacag	4560
acaagctgtg	accgtctccg	ggagctgcat	gtgtcagagg	ttttcaccgt	catcacccga	4620
acgcgcgaga	cgaaggggcc	tcgtgatagc	cctattttta	taggttaatg	tcatgataat	4680
aatggtttct	tagacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	ccccatttg	4740
tttatttttc	taaatacatt	caaataatgta	tccgctcatg	agacaataac	cctgataaat	4800
gcttcaataa	tattgaaaaa	ggaagagtat	gagtattcaa	catttccgtg	tcgcccttat	4860
tccctttttt	cgggcatttt	gccttcctgt	ttttgctcac	ccagaaacgc	tgggtgaaagt	4920
aaaagatgct	gaagatcagt	tgggtgcacg	agtgggttac	atcgaactgg	atctcaacag	4980
cggtaagatc	cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	5040
agttctgcta	tgtggcgcgg	tattatcccg	tattgacgcc	gggcaagagc	aactcggctg	5100
ccgcatacac	tattctcaga	atgacttggt	tgagtactca	ccagtccacag	aaaagcatct	5160
tacggatggc	atgacagtaa	gagaattatg	cagtgtgcc	ataaccatga	gtgataaac	5220
tgccggccaac	ttactcttga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	5280
caacatgggg	gatcatgtaa	ctcgcttga	tcgttgggaa	ccggagctga	atgaagccat	5340
accaaacgac	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgccgaaact	5400
attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	5460
ggataaagtt	gcaggaccac	ttctgcgtc	ggcccttcgg	gctggctggg	ttattgctga	5520
taaatctgga	gccggtgagc	gtgggtctcg	cggatcatt	gcagcactgg	ggccagatgg	5580
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggatgaacg	5640
aaatagacag	atcgtcgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	5700
agtttactca	tataactttt	agattgattt	aaaacttcac	ttttaattta	aaaggatcta	5760
ggtgaagatc	ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	5820
ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctcgg	5880
cgtaatctgc	tgcttgcaaa	caaaaaaac	accgctacca	gcggtgggtt	gtttgccgga	5940
tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	6000
tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	6060
tacatacctc	gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtctg	6120
tcttaccggg	ttggactcaa	gacgatagtt	accggataag	gcgcagcggg	cgggctgaac	6180
gggggggttcg	tgcacacagc	ccagcttggg	gcgaacgacc	tacaccgaac	tgagatacct	6240
acagcgtgag	ctatgagaaa	gcgccacgct	tcccgaaggg	agaaaaggcg	acaggtatcc	6300
ggtaagcggc	agggtcggaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	6360
gtatctttat	agtctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	6420
ctcgtcaggg	gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	6480
ggccttttgc	tggccttttg	ctcacatggc	tcgacagatc	t		6521

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097059 A3(51) International Patent Classification: C07H 21/04,
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier applications:US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)(71) Applicant (for all designated States except US): CHRO-
MOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081
Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward
[US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6
(CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Ave-
nue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM,
Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C.
V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence
Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine
[CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8
(CA). FLEMING, Elena [CA/CA]; 248 E 18th, North
Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra
[CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3
(CA). SHELLARD, Joan [CA/CA]; #215-1345 West
15th Avenue, Vancouver, B.C. V6H 3R3 (CA).(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman
White & McAuliffe LLP, 7th floor, 4350 La Jolla Village
Drive, San Diego, CA 92122-1246 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for the following design-
ations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG)— as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for the following design-
ations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG)

Published:

— with international search report

(88) Date of publication of the international search report:
30 May 2003

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, re-
combination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational
engineering of the chromosome for a variety of applications.

WO 02/097059 A3



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/85,87,90; A01N 43/34; C12N 15/09

US CL : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS (EAST); STN (MEDLINE, BIOSIS, CAPLUS)**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 01/07572 A2 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 01 February 2001 (01.02.2001), see entire document.	1,2,7,8,11,23-25,27,50 ----- 3-6,9,10,12,13,26,28-
X --- Y	WO 00/11155 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 02 March 2000 (02.03.2000), see entire document.	1,2,7,8,11,23-25,27,50 ----- 3-6,9,10,12,13,26,28-
X --- Y	US 6,171,861 B1 (HARTLEY et al.) 09 January 2001 (09.01.2001), see entire document, especially sequence listing.	1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85 ----- 4-7,9,13-22,28-30,32,34,35,37,38,41,42,46-49,51-53,56,67-71,84,86



Further documents are listed in the continuation of Box C.



See patent family annex.

Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"Z" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

01 November 2002 (01.11.2002)

Date of mailing of the international search report

27 NOV 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Daniel M Sullivan

Telephone No. 703-308-0196

INTERNATIONAL SEARCH REPORT

PCT/US02/17452

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 96/40724 A1 (LIFE TECHNOLOGIES, INC.) 19 December 1996 (19.12.96), see entire document, especially SEQ ID NO: 1-16.	1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85 ----- 1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85
Y	WO 94/00569 A1 (GENPHARM INTERNATIONAL, INC.) 06 January 1994 (06.01.94), see entire document.	72-78
Y	WO 94/23049 A2 (THE JOHNS HOPKINS UNIVERSITY) 13 October 1994 (13.10.94), see entire document.	72-78
X --- Y	US 5,721,118 A (SCHEFFLER) 24 February 1998 (24.02.98), see entire document, especially Example 3.	14,15,22-25,57 ----- 16-21,26,58-64,72-78,84,85,98-105,123
X --- Y	US 5,948,653 A (PATI et al.) 07 September 1999 (07.09.99), see entire document, especially the first full paragraph in column 29.	14,15,22-25,57 ----- 16-21,26,58-64,72-78,84,85,98-105,123
X --- Y	GB 2 331 752 A (MEDICAL RESEARCH COUNCIL) 02 June 1999 (02.06.99), see entire document, especially pages 8-10.	109,123 ----- 72-78,110-122
A	HALDIMANN et al. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol. November 2001, Vol. 183 No. 21, pages 6384-6393, see entire document.	65,66
X,P --- Y,P	MORALLI et al. Insertion of a loxP site in a size-reduced human accessory chromosome. Cytogenet. Cell Genet. 2001, Vol. 94 No. 3-4, pages 113-120, see entire document.	91-93,106,108 ----- 94-97,107
X --- Y	LORBACH et al. Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. J. Mol. Biol. March 2000, Vol. 296 No. 5, pages 1175-1181, see entire document.	1,2,8,50 ----- 3-7,9-13,23-32,51-56,84,85
X --- Y	CALL et al. A cre-lox recombination system for the targeted integration of circular yeast artificial chromosomes into embryonic stem cells. Hum. Mol. Genet. July 2000, Vol. 9, No. 12, pages 1745-1751, see entire document.	14,15,22-26,57,59 ----- 16-21,58,60,84,85,98-105
Y,P	HADLACZKY, G. Satellite DNA-based artificial chromosomes for use in gene therapy. Curr. Opin. Mol. Ther. April 2001, Vol. 3 No. 2, pages 125-132, see entire document.	1-123

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/17452

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-64, 67-71, 79, 84-96, 91-108, and 123, drawn to eukaryotic recombinogenic chromosomes used for introducing heterologous nucleic acids into a chromosome and the resulting cells.

Group II, claim(s) 65-66 and 87-88, drawn to a lambda intR mutein.

Group III, claim(s) 72-78, drawn to the production of transgenic animals.

Group IV, claim(s) 80, drawn to the production of an artificial chromosome library.

Group V, claim(s) 81-83, drawn to a library of cells for genomic screening.

Group VI, claim(s) 89-90, drawn to a modified iron-induced promoter.

Group VII, claim(s) 109-122, drawn to a method for screening compounds and their effects on regulatory regions.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of Group I which defines an advance over the art is a eukaryotic chromosome containing recombinogenic sites that can be used to introduce heterologous nucleic acids into chromosomes, and cells containing these recombinogenic chromosomes.

The special technical feature of Group II involves a lambda intR mutein. This feature defines an advance over Group I in that it involves a protein that is not required for the technical features as set forth above in Group I.

The special technical feature of Group III involves the production of a transgenic animal, which represents a second method for using the invention as set forth above in Group I.

The special technical feature of Group IV involves the production of artificial chromosome expression system libraries, which represents a third method for using the invention as set forth above in Group I.

The special technical feature of Group V involves a library of cells containing the artificial chromosome expression system libraries set forth above in Group IV, and represents a first product resulting from Group IV.

The special technical feature of Group VI involves a modified iron-inducible promoter. This feature defines an advance over Group I in that it involves a promoter that is not required for the technical features as set forth above in Group I.

The special technical feature of Group VII involves a method for screening compounds for their effect on regulatory regions, which represents a fourth method of using the invention as set forth above in Group I.

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 2002/097059 A3

(51) International Patent Classification⁷: **C07H 21/04**,
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number:
PCT/US2002/017452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier applications:

US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): **CHROMOS MOLECULAR SYSTEMS, INC.** [CA/CA]; 8081
Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PERKINS, Edward**
[US/US]; 4203 Robinson Street, Duluth, MN 55804 (US). **PEREZ, Carl** [US/CA]; 95 West 11th Avenue, Vancouver,
British Columbia V5Y 1S6 (CA). **LINDENBAUM, Michael** [CA/CA]; 9941 Martin Court, Burnaby, British
Columbia V3K 1J5 (CA). **GREENE, Amy** [US/US];
4203 Robinson Street, Duluth, MN 55804 (US). **LEUNG, Josephine** [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C.
V3J 7P8 (CA). **FLEMING, Elena** [CA/CA]; 427 Montroyal Blvd., North Vancouver, British Columbia V7N
3E2 (CA). **STEWART, Sandra** [CA/CA]; 2618 Oxford
Street, Vancouver, B.C. V5K 1N3 (CA). **SHELLARD, Joan** [CA/CA]; #215-1345 West 15th Avenue, Vancouver,
B.C. V6H 3R3 (CA).

(74) Agents: **SEIDMAN, Stephanie, L.** et al.; Heller Ehrman
White & McAuliffe LLP, 7th floor, 4350 La Jolla Village
Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for the following design-
ations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG)
- as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for the following design-
ations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG)

Published:

- with international search report
- with amended claims

(88) Date of publication of the International search report:
30 May 2003

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, re-
combination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational
engineering of the chromosome for a variety of applications.

WO 2002/097059 A3



Date of publication of the amended claims: 31 December 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(15) Information about Correction:

Previous Correction:

see PCT Gazette No. 39/2003 of 25 September 2003, Section II

AMENDED CLAIMS

[received by the International Bureau on 27 January 2003 (27.01.2003);
Original claims 22, 29, 39, 40, 68, 104, 107, 108, 114 and 121 replaced by
Amended Claims 22, 29, 39, 40, 68, 104, 107, 108, 114 and 121.;
Remaining claims unchanged]

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).
14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase
5 catalyzed recombination.
15. The *ACes* of claim 14 that contains one site.
16. The *ACes* of claim 14 that is predominantly heterochromatin.
17. The *ACes* of claim 14 that contains no more than about
30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
- 10 18. The *ACes* of claim 14 that is a plant *ACes*.
19. The *ACes* of claim 14 that is an animal *ACes*.
20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.
21. The *ACes* of claim 14 that is a fish *ACes*.
- 15 22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli*, the R/RS
20 system of the pSR1 plasmid, or any combination thereof.
23. A method of introducing heterologous nucleic acid into a chromosome, comprising:
contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a
25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

28. The combination of claim 27, further comprising nucleic acid
15 encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase
20 system of phage Mu, the Pin recombinase system of *E. coli*, the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

31. The combination of claim 27, wherein a vector is the plasmid
25 pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

- (a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second recombination site and the heterologous nucleic acid;
- (b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis, Cre, $\gamma\delta$, Tn3 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 33, wherein the recombination sites are selected from the group consisting of att and lox P sites.

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence
10 of an IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

70. The method of claim 69, wherein the selectable marker is
15 selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an
25 embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the PAC comprises transcriptional regulatory sequence of nucleotides derived from animals.

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:

- 5 introducing a vector of claim 91 into a cell;
 growing the cells; and
 selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein a sufficient portion of the
10 vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:

- 15 contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

 the reporter *ACes* comprises one or a plurality of reporter constructs;

- a reporter construct comprises a reporter gene in operative linkage
20 with a regulatory region responsive to test or known compounds; and
 detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively
25 linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent
5 proteins.

113. The method of claim 109, wherein the *ACes* comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cells.

122. The method of claim 109, wherein the cell is in an animal;
and the method comprises whole-body imaging to monitor expression of
5 the reporter in the animal.

123. A reporter *ACes* comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
5 December 2002 (05.12.2002)(10) International Publication Number
PCT WO 02/097059 A3

- (51) International Patent Classification⁷: C07H 21/04, C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09
- (21) International Application Number: PCT/US02/17452
- (22) International Filing Date: 30 May 2002 (30.05.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/294,758 30 May 2001 (30.05.2001) US
60/366,891 21 March 2002 (21.03.2002) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/294,758 (CIP)
Filed on 30 May 2001 (30.05.2001)
US 60/366,891 (CIP)
Filed on 21 March 2002 (21.03.2002)
- (71) Applicant (for all designated States except US): CHROMOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081 Loughheed Highway, Burnaby, B.C. V5A 1W9 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): PERKINS, Edward [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). FLEMING, Elena [CA/CA]; 248 E 18th, North Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). SHELLARD, Joan [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).
- (74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Declarations under Rule 4.17:**
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.



WO 02/097059 A3



GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

(48) Date of publication of this corrected version:

25 September 2003

Published:

— with international search report

(15) Information about Correction:

see PCT Gazette No. 39/2003 of 25 September 2003, Section II

(88) Date of publication of the international search report:

30 May 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

-1-

CHROMOSOME-BASED PLATFORMS**RELATED APPLICATIONS**

Benefit of priority to U.S. provisional application Serial No. 60/294,758, filed May 30, 2001, to Perkins, *et al.*, entitled "CHROMOSOME-BASED PLATFORMS" and to U.S. provisional application Serial No. 60/366,891, filed March 21, 2002, to Perkins, *et al.*, entitled

5 "CHROMOSOME-BASED PLATFORMS" is claimed. Where permitted, the subject matter of which are herein incorporated by reference in their entirety.

This application is related to Provisional Application No. 60/294,687, filed May 30, 2001, by CARL PEREZ AND STEVEN

10 FABIJANSKI entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES* and to U.S. Provisional Application No. 60/296,329, filed June 4, 2001, by CARL PEREZ AND STEVEN FABIJANSKI entitled

15 *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES*. This application also is related to U.S. Provisional Application No. 60/294,758, filed May 30, 2001, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS* and to U.S. Provisional Application No. 60/366,891, filed

20 March 21, 2002, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS*. This application is also related to U.S. application Serial Nos. (attorney dkt nos. 24601-419 and 419PC), filed on the same day herewith, entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS OF PREPARING PLANT ARTIFICIAL CHROMOSOMES* to Perez *et al.*

25 This application is related to U.S. application Serial No. 08/695,191, filed August 7, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*

METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES, now U.S. Patent No. 6,025,155. This application is also related to U.S. application Serial No. 08/682,080, filed July 15, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF*
5 *AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,077,697. This application is also related U.S. application Serial No. 08/629,822, filed April 10, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING*
10 *ARTIFICIAL CHROMOSOMES* (now abandoned), and is also related to copending U.S. application Serial No. 09/096,648, filed June 12, 1998, by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* and to U.S. application Serial No.
15 09/835,682, April 10, 1997 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned). This application is also related to copending U.S. application Serial No. 09/724,726, filed November 28, 2000, U.S. application Serial
20 No. 09/724,872, filed November 28, 2000, U.S. application Serial No. 09/724,693, filed November 28, 2000, U.S. application Serial No. 09/799,462, filed March 5, 2001, U.S. application Serial No. 09/836,911, filed April 17, 2001, and U.S. application Serial No. 10/125,767, filed April 17, 2002, each of which is by GYULA
25 HADLACZKY and ALADAR SZALAY, and is entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also related to International PCT application No. WO 97/40183. Where permitted the

subject matter of each of these provisional applications, international applications, and applications is incorporated by reference in its entirety.

FIELD OF INVENTION

Artificial chromosomes, including *ACes*, that have been engineered
5 to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome.

BACKGROUND

Artificial chromosomes

10 A variety of artificial chromosomes for use in plants and animals, particularly higher plants and animals are available. In particular, U.S. Patent Nos. 6,025,155 and 6,077,697 provide heterochromatic artificial chromosomes designated therein as satellite artificial chromosomes (SATACs) and now designated artificial chromosome expression systems
15 (*ACes*). These chromosomes are prepared by introducing heterologous DNA into a selected plant or animal cell under conditions that result in integration into a region of the chromosome that leads to an amplification event resulting in production of a dicentric chromosome. Subsequent treatment and growth of cells with dicentric chromosomes, including
20 further amplifications, ultimately results in the artificial chromosomes provided therein. In order to introduce a desired heterologous gene (or a plurality of heterologous genes) into the artificial chromosome, the process is repeated introducing the desired heterologous genes and nucleic acids in the initial targeting step. This process is time consuming
25 and tedious. Hence, more tractable and efficient methods for introducing heterologous nucleic acid molecules into artificial chromosomes, particularly *ACes*, are needed.

Therefore, it is an object herein to provide engineered artificial chromosomes that permit tractable, efficient and rational engineering of artificial chromosomes.

SUMMARY OF THE INVENTION

5 Provided herein are artificial chromosomes that permit tractable, efficient and rational engineering thereof. In particular, the artificial chromosomes provided herein contain one or a plurality of loci (sites) for site-specific, recombination-directed integration of DNA. Thus, provided herein are platform artificial chromosome expression systems ("platform
10 *ACes*") containing single or multiple site-specific, recombination sites. The artificial chromosomes and *ACes* artificial chromosomes include plant and animal chromosomes. Any recombinase system that effects site-specific recombination is contemplated for use herein.

 In one embodiment, chromosomes, including platform *ACes*, are
15 provided that contain one or more lambda *att* sites designed for recombination-directed integration in the presence of lambda integrase, and that are mutated so that they do not require additional factors. Methods for preparing such chromosomes, vectors for use in the methods, and uses of the resulting chromosomes are also provided.

20 Platform *ACes* containing the recombination site(s) and methods for introducing heterologous nucleic acid into such sites and vectors therefor, are provided.

 Also provided herein is a bacteriophage lambda (λ) integrase site-specific recombination system.

25 Methods using recombinase mediated recombination target gene expression vectors and/or genes for insertion thereof into platform chromosomes and the resulting chromosomes are provided.

 Combinations and kits containing the combinations of vectors encoding a recombinase and integrase and primers for introduction of the

site recognized thereby are also provided. The kits optionally include instructions for performing site-directed integration or preparation of *ACes* containing such sites.

Also provided herein are mammalian and plant cells comprising the
5 artificial chromosomes and *ACes* described herein. The cells can be nuclear donor cells, stem cells, such as a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell or an embryonic stem cell.

Also provided is a lambda-intR mutain comprising a glutamic acid to arginine change at position 174 of wild-type lambda-integrase3. Also
10 provided are transgenic animals and methods for producing a transgenic animal, comprising introducing a *ACes* into an embryonic cell, such as a stem cell or embryo. The *ACes* can comprise heterologous nucleic acid that encodes a therapeutic product. The transgenic animal can be a fish, insect, reptile, amphibians, arachnid or mammal. In certain embodiments,
15 the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

The platform *ACes*, including plant and animal *ACes*, such as MACs, provided herein can be introduced into cells, such as, but not
20 limited to, animal cells, including mammalian cells, and into plant cells. Hence plant cells that contain platform MACs, animal cells that contain platform PACs and other combinations of cells and platform *ACes* are provided.

DESCRIPTION OF FIGURES

25 **FIGURE 1** provides a diagram depicting creation of an exemplary *ACes* artificial chromosome prepared using methods detailed in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183. In this exemplified embodiment, the nucleic acid is targeted to an acrocentric chromosome in an animal or plant, and the

heterologous nucleic acid includes a sequence-specific recombination site and marker genes.

FIGURE 2 provides a map of pWEPuro9K, which is a targeting vector derived from the vector pWE15 (GenBank Accession # X65279; SEQ ID No. 31). Plasmid pWE15 was modified by replacing the *Sa*I (Klenow filled)/*Sma*I neomycin resistance encoding fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance-encoding fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc., Palo Alto, CA; GenBank Accession no. U07648; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (see Example 1, see, also Csonka *et al.* (2000) *Journal of Cell Science* 113:3207-32161; and SEQ ID NO: 118), containing a portion of the mouse rDNA region, was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K.

FIGURE 3 depicts construction of an ACes platform chromosome with a single recombination site, such as loxP sites or an *attP* or *attB* site. This platform ACes chromosome is an exemplary artificial chromosome with a single recombination site.

FIGURE 4 provides a map of plasmid pSV40-193attPsensePur.

FIGURE 5 depicts a method for formation of a chromosome platform with multiple recombination integration sites, such as *attP* sites.

FIGURE 6 sets forth the sequences of the core region of *attP*, *attB*, *attL* and *attR* (SEQ ID Nos. 33-36).

FIGURE 7 depicts insertional recombination of a vector encoding a marker gene, DsRed and an *attB* site with an artificial chromosome containing an *attP* site.

FIGURE 8 provides a map of plasmid pCXLamIntR (SEQ ID NO: 112), which includes the Lambda integrase (E174R)-encoding nucleic acid.

FIGURE 9 diagrammatically summarizes the platform technology; marker 1 permits selection of the artificial chromosomes containing the integration site; marker 2, which is promoterless in the target gene expression vector, permits selection of recombinants. Upon
5 recombination with the platform marker 2 is expressed under the control of a promoter resident on the platform.

FIGURE 10 provides the vector map for the plasmid p18attBZEO-5'6XHS4eGFP (SEQ ID NO: 116).

FIGURE 11 provides the vector map for the plasmid p18attBZEO-
10 3'6XHS4eGFP (SEQ ID NO: 115).

FIGURE 12 provides the vector map for the plasmid p18attBZEO-(6XHS4)2eGFP (SEQ ID NO: 110).

FIGURES 13 AND 14 depict the integration of a PCR product by site-specific recombination as set forth in Example 8.

15 FIGURE 15 provides the vector map for the plasmid pPACrDNA as set forth in Example 9.A.

DETAILED DESCRIPTION OF THE INVENTION

A. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used
20 herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference
25 in their entirety. Where reference is made to a URL or other such identifier or address, it understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

As used herein, nucleic acid refers to single-stranded and/or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique and of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleotides long.

As used herein, DNA is meant to include all types and sizes of DNA molecules including cDNA, plasmids and DNA including modified nucleotides and nucleotide analogs.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified-nucleotides, such as, but are not limited to, phosphorothioate nucleotides and deazapurine nucleotides and other nucleotide analogs.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations and/or in amounts in a genome or cell that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not

normally produced by the cell in which it is expressed. Any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Heterologous DNA and RNA may also encode RNA or
5 proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes.

Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of modification of the endogenous genes or for production of an
10 encoded protein. For example, a heterologous or foreign gene may be isolated from a different species than that of the host genome, or alternatively, may be isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Other examples of heterologous DNA include, but
15 are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers traits including, but not limited to, herbicide, insect, or disease resistance; traits, including, but not limited to, oil quality or carbohydrate composition. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the
20 heterologous DNA has been introduced.

As used herein, operative linkage or operative association, or grammatical variations thereof, of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences
25 refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the

promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA.

In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak (1991) *J. Biol. Chem.* 266:19867-19870) can be inserted immediately 5' of the start codon and may enhance expression.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, preferably under moderate or high stringency conditions, forming a stable duplex. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. The longer the hybridizing nucleic acid, the more base mismatches it can contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

As used herein, regulatory molecule refers to a polymer of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or a polypeptide that is capable of enhancing or inhibiting expression of a gene.

As used herein, recognition sequences are particular sequences of nucleotides that a protein, DNA, or RNA molecule, or combinations thereof, (such as, but not limited to, a restriction endonuclease, a modification methylase and a recombinase) recognizes and binds. For example, a recognition sequence for Cre recombinase (see, *e.g.*, SEQ ID NO:58) is a 34 base pair sequence containing two 13 base pair inverted

repeats (serving as the recombinase binding sites) flanking an 8 base pair core and designated loxP (see, *e.g.*, Sauer (1994) *Current Opinion in Biotechnology* 5:521-527). Other examples of recognition sequences, include, but are not limited to, attB and attP, attR and attL and others
5 (see, *e.g.*, SEQ ID Nos. 8, 41-56 and 72), that are recognized by the recombinase enzyme Integrase (see, SEQ ID Nos. 37 and 38 for the nucleotide and encoded amino acid sequences of an exemplary lambda phage integrase).

The recombination site designated attB is an approximately 33 base
10 pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region; attP (SEQ ID No. 72) is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis (see, *e.g.*, Landy (1993) *Current Opinion in Biotechnology* 3:699-707) see,
15 *e.g.*, SEQ ID Nos. 8 and 72).

As used herein, a recombinase is an enzyme that catalyzes the exchange of DNA segments at specific recombination sites. An integrase herein refers to a recombinase that is a member of the lambda (λ) integrase family.

20 As used herein, recombination proteins include excisive proteins, integrative proteins, enzymes, co-factors and associated proteins that are involved in recombination reactions using one or more recombination sites (see, Landy (1993) *Current Opinion in Biotechnology* 3:699-707). The recombination proteins used herein can be delivered to a cell via an
25 expression cassette on an appropriate vector, such as a plasmid, and the like. In other embodiments, the recombination proteins can be delivered to a cell in protein form in the same reaction mixture used to deliver the desired nucleic acid, such as a platform ACes, donor target vectors, and the like.

As used herein the expression "lox site" means a sequence of nucleotides at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination event. A LoxP site is a 34 base pair nucleotide sequence from bacteriophage P1 (see, 5 *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402). The LoxP site contains two 13 base pair inverted repeats separated by an 8 base pair spacer region as follows: (SEQ ID NO. 57):

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT

E. coli DH5Δlac and yeast strain BSY23 transformed with plasmid pBS44 10 carrying two loxP sites connected with a LEU2 gene are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53254 and ATCC 20773, respectively. The lox sites can be isolated from plasmid pBS44 with restriction enzymes *EcoRI* and *Sall*, or *XhoI* and *BamHI*. In addition, a preselected DNA segment can be inserted 15 into pBS44 at either the *Sall* or *BamHI* restriction enzyme sites. Other lox sites include, but are not limited to, LoxB, LoxL, LoxC2 and LoxR sites, which are nucleotide sequences isolated from *E. coli* (see, *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398). Lox sites can also be produced by a variety of synthetic techniques (see, *e.g.*, Ito *et al.* (1982) 20 *Nuc. Acid Res.* 10:1755 and Ogilvie *et al.* (1981) *Science* 270:270).

As used herein, the expression "cre gene" means a sequence of nucleotides that encodes a gene product that effects site-specific recombination of DNA in eukaryotic cells at lox sites. One cre gene can be isolated from bacteriophage P1 (see, *e.g.*, Abremski *et al.* (1983) *Cell* 25 32:1301-1311). *E. coli* DH1 and yeast strain BSY90 transformed with plasmid pBS39 carrying a cre gene isolated from bacteriophage P1 and a GAL1 regulatory nucleotide sequence are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53255

and ATCC 20772, respectively. The cre gene can be isolated from plasmid pBS39 with restriction enzymes *XhoI* and *Sall*.

As used herein, site-specific recombination refers to site-specific recombination that is effected between two specific sites on a single
5 nucleic acid molecule or between two different molecules that requires the presence of an exogenous protein, such as an integrase or recombinase.

For example, Cre-lox site-specific recombination can include the following three events:

- 10 a. deletion of a pre-selected DNA segment flanked by lox sites;
- b. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites; and
- c. reciprocal exchange of DNA segments proximate to
15 lox sites located on different DNA molecules.

This reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. DNA segment refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source.

20 Since the lox site is an asymmetrical nucleotide sequence, two lox sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA segment located between the two lox sites and a connection between the resulting ends of the original
25 DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox

sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the gene product of the cre gene. Thus, the Cre-lox system can be used to specifically delete, invert, or
5 insert DNA. The precise event is controlled by the orientation of lox DNA sequences, in *cis* the lox sequences direct the *Cre* recombinase to either delete (lox sequences in direct orientation) or invert (lox sequences in inverted orientation) DNA flanked by the sequences, while *in trans* the lox sequences can direct a homologous recombination event resulting in the
10 insertion of a recombinant DNA.

As used herein, a chromosome is a nucleic acid molecule, and associated proteins, that is capable of replication and segregation within a cell upon cell division. Typically, a chromosome contains a centromeric region, replication origins, telomeric regions and a region of nucleic acid
15 between the centromeric and telomeric regions.

As used herein, a centromere is any nucleic acid sequence that confers an ability to segregate to daughter cells through cell division. A centromere may confer stable segregation of a nucleic acid sequence, including an artificial chromosome containing the centromere, through
20 mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A particular centromere is not necessarily derived from the same species in which it is introduced, but has the ability to promote DNA segregation in cells of that species.

As used herein, euchromatin and heterochromatin have their
25 recognized meanings. Euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to be transcriptionally inactive. Highly repetitive DNA sequences (satellite DNA) are usually located in regions of the heterochromatin surrounding

the centromere (pericentric or pericentromeric heterochromatin).

Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

5 As used herein, an acrocentric chromosome refers to a chromosome with arms of unequal length.

 As used herein, endogenous chromosomes refer to genomic chromosomes as found in a cell prior to generation or introduction of an artificial chromosome.

10 As used herein, artificial chromosomes are nucleic acid molecules, typically DNA, that stably replicate and segregate alongside endogenous chromosomes in cells and have the capacity to accommodate and express heterologous genes contained therein. It has the capacity to act as a gene delivery vehicle by accommodating and expressing foreign genes
15 contained therein. A mammalian artificial chromosome (MAC) refers to chromosomes that have an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes that include centromeres that function in plant, insect and avian cells, respectively. A human artificial
20 chromosome (HAC) refers to chromosomes that include centromeres that function in human cells. For exemplary artificial chromosomes, see, *e.g.*, U.S. Patent Nos. 6,025,155; 6,077,697; 5,288,625; 5,712,134; 5,695,967; 5,869,294; 5,891,691 and 5,721,118 and published International PCT application Nos, WO 97/40183 and WO 98/08964.
25 Artificial chromosomes include those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), minichromosomes that contain a *de novo* centromere (see, U.S. Patent

Nos. 5,712,134, 5,891,691 and 5,288,625), artificial chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA and *in vitro* assembled artificial chromosomes (see, copending U.S. 5 provisional application Serial No. 60/294,687, filed on May 30, 2001).

As used herein, the term "satellite DNA-based artificial chromosome (SATAC)" is interchangeable with the term "artificial chromosome expression system (ACes)". These artificial chromosomes (ACes) include those that are substantially all neutral non-coding 10 sequences (heterochromatin) except for foreign heterologous, typically gene-encoding nucleic acid, that is interspersed within the heterochromatin for the expression therein (see U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183), or that is in a single locus as provided herein. Also included 15 are ACes that may include euchromatin and that result from the process described in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183 and outlined herein. The delineating structural feature is the presence of repeating units, that are generally predominantly heterochromatin. The precise structure of the ACes will 20 depend upon the structure of the chromosome in which the initial amplification event occurs; all share the common feature of including a defined pattern of repeating units. Generally ACes have more heterochromatin than euchromatin. Foreign nucleic acid molecules (heterologous genes) contained in these artificial chromosome expression 25 systems can include any nucleic acid whose expression is of interest in a particular host cell. Such foreign nucleic acid molecules, include, but are not limited to; nucleic acid that encodes traceable marker proteins (reporter genes), such as fluorescent proteins, such as green, blue or red fluorescent proteins (GFP, BFP and RFP, respectively), other reporter

genes, such as β -galactosidase and proteins that confer drug resistance, such as a gene encoding hygromycin-resistance. Other examples of heterologous nucleic acid molecules include, but are not limited to, DNA that encodes therapeutically effective substances, such as anti-cancer
5 agents, enzymes and hormones, DNA that encodes other types of proteins, such as antibodies, and DNA that encodes RNA molecules (such as antisense or siRNA molecules) that are not translated into proteins.

As used herein, an artificial chromosome platform, also referred to
10 herein as a "platform ACes" or "ACes platform", refers to an artificial chromosome that has been engineered to include one or more sites for site-specific, recombination-directed integration. In particular, ACes that are so-engineered are provided. Any sites, including but not limited to any described herein, that are suitable for such integration are
15 contemplated. Plant and animal platform ACes are provided. Among the ACes contemplated herein are those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), artificial
20 chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA resulting from an amplification event depicted in the referenced patent and herein. Included among the ACes for use in generating platforms, are artificial chromosomes that introduce and
25 express heterologous nucleic acids in plants (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001). These include artificial chromosomes that have a centromere derived from a plant, and, also, artificial chromosomes that have centromeres that may be derived from other organisms but that function in plants.

As used herein a "reporter ACes" refers to an ACes that comprises one or a plurality of reporter constructs, where the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

5 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of substantially similar or identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

10 As used herein, amplification-based artificial chromosomes are artificial chromosomes derived from natural or endogenous chromosomes by virtue of an amplification event, such as one initiated by introduction of heterologous nucleic acid into rDNA in a chromosome. As a result of such an event, chromosomes and fragments thereof exhibiting segmented
15 or repeating patterns arise. Artificial chromosomes can be formed from these chromosomes and fragments. Hence, amplification-based artificial chromosomes refer to engineered chromosomes that exhibit an ordered segmentation that is not observed in naturally occurring chromosomes and that distinguishes them from naturally occurring chromosomes. The
20 segmentation, which can be visualized using a variety of chromosome analysis techniques known to those of skill in the art, correlates with the structure of these artificial chromosomes. In addition to containing one or more centromeres, the amplification-based artificial chromosomes, throughout the region or regions of segmentation are predominantly made
25 up of nucleic acid units also referred to as "amplicons", that is (are) repeated in the region and that have a similar gross structure. Repeats of an amplicon tend to be of similar size and share some common nucleic acid sequences. For example, each repeat of an amplicon may contain a replication site involved in amplification of chromosome segments and/or

some heterologous nucleic acid that was utilized in the initial production of the artificial chromosome. Typically, the repeating units are substantially similar in nucleic acid composition and may be nearly identical.

5 The amplification-based artificial chromosomes differ depending on the chromosomal region that has undergone amplification in the process of artificial chromosome formation. The structures of the resulting chromosomes can vary depending upon the initiating event and/or the conditions under which the heterologous nucleic acid is introduced,
10 including modification to the endogenous chromosomes. For example, in some of the artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of heterochromatic DNA. In other artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of euchromatic DNA or may
15 be made up of similar amounts of heterochromatic and euchromatic DNA.

As used herein an amplicon is a repeated nucleic acid unit. In some of the artificial chromosomes described herein, an amplicon may contain a set of inverted repeats of a megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to
20 some of the predominantly heterochromatic artificial chromosomes, the megareplicon can contain a set of tandem DNA blocks (*e.g.*, ~7.5 Mb DNA blocks) each containing satellite DNA flanked by non-satellite DNA or may be made up of substantially rDNA. Contained within the megareplicon is a primary replication site, referred to as the
25 megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller (*e.g.*, 50-300 kb) secondary replicons.

In artificial chromosomes, such as those provided U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183, the megareplicon is defined by two tandem blocks (~7.5 Mb DNA blocks in the chromosomes provided therein). Within each artificial
5 chromosome or among a population thereof, each amplicon has the same gross structure but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the artificial chromosomes or their overall
10 structure as described herein.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating artificial chromosomes provided herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other
15 cellular events involving recombination (*e.g.*, DNA repair). Such regions include regions of the chromosome that contain tandem repeats, such as satellite DNA, rDNA, and other such sequences.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more
20 than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments is a replicable
25 chromosome. If one of the chromosomes undergoes amplification of primarily euchromatic DNA to produce a fully functional chromosome that is predominantly (at least more than 50%) euchromatin, it is a minichromosome. The remaining chromosome is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification,

whereby heterochromatin (such as, for example, satellite DNA) is amplified and a euchromatic portion (such as, for example, an arm) remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a predominantly heterochromatic artificial chromosome. Predominantly heterochromatic artificial chromosomes can be produced from other partially heterochromatic artificial chromosomes by culturing the cell containing such chromosomes under conditions such as BrdU treatment that destabilize the chromosome and/or growth under selective conditions so that a predominantly heterochromatic artificial chromosome is produced. For purposes herein, it is understood that the artificial chromosomes may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous DNA. Typically, artificial chromosomes appear after about 5 to about 60, or about 5 to about 55, or about 10 to about 55 or about 25 to about 55 or about 35 to about 55 cell doublings after initiation of artificial chromosome generation, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, an artificial chromosome that is predominantly heterochromatic (*i.e.*, containing more heterochromatin than euchromatin, typically more than about 50%, more than about 70%, or more than about 90% heterochromatin) may be produced by introducing nucleic acid molecules into cells, such as, for example, animal or plant cells, and selecting cells that contain a predominantly heterochromatic artificial chromosome. Any nucleic acid may be introduced into cells in such methods of producing the artificial chromosomes. For example, the nucleic acid may contain a selectable marker and/or optionally a sequence that targets nucleic acid to the pericentric, heterochromatic region of a chromosome, such as in the short arm of acrocentric chromosomes and

nucleolar organizing regions. Targeting sequences include, but are not limited to, lambda phage DNA and rDNA for production of predominantly heterochromatic artificial chromosomes in eukaryotic cells.

After introducing the nucleic acid into cells, a cell containing a
5 predominantly heterochromatic artificial chromosome is selected. Such
cells may be identified using a variety of procedures. For example,
repeating units of heterochromatic DNA of these chromosomes may be
discerned by G-banding and/or fluorescence in situ hybridization (FISH)
techniques. Prior to such analyses, the cells to be analyzed may be
10 enriched with artificial chromosome-containing cells by sorting the cells
on the basis of the presence of a selectable marker, such as a reporter
protein, or by growing (culturing) the cells under selective conditions. It
is also possible, after introduction of nucleic acids into cells, to select
cells that have a multicentric, typically dicentric, chromosome, a formerly
15 multicentric (typically dicentric) chromosome and/or various
heterochromatic structures, such as a megachromosome and a sausage
chromosome, that contain a centromere and are predominantly
heterochromatic and to treat them such that desired artificial
chromosomes are produced. Cells containing a new chromosome are
20 selected. Conditions for generation of a desired structure include, but are
not limited to, further growth under selective conditions, introduction of
additional nucleic acid molecules and/or growth under selective conditions
and treatment with destabilizing agents, and other such methods (see
International PCT application No. WO 97/40183 and U.S. Patent Nos.
25 6,025,155 and 6,077,697).

As used herein, a "selectable marker" is a nucleic acid segment,
generally DNA, that allows one to select for or against a molecule or a
cell that contains it, often under particular conditions. These markers can
encode an activity, such as, but not limited to, production of RNA,

peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds and compositions. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products that provide resistance against otherwise
5 toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be identified, such as phenotypic markers,
10 including β -galactosidase, red, blue and/or green fluorescent proteins (FPs), and cell surface proteins; (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense
15 oligonucleotides or siRNA molecules for use in RNA interference); (7) nucleic acid segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) nucleic acid segments that can be used to isolate a desired molecule (e.g. specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be
20 otherwise non-functional, such as for PCR amplification of subpopulations of molecules; and/or (10) nucleic acid segments, which when absent, directly or indirectly confer sensitivity to particular compounds. Thus, for example, selectable markers include nucleic acids encoding fluorescent proteins, such as green fluorescent proteins, β -galactosidase and other
25 readily detectable proteins, such as chromogenic proteins or proteins capable of being bound by an antibody and FACs sorted. Selectable markers such as these, which are not required for cell survival and/or proliferation in the presence of a selection agent, are also referred to herein as reporter molecules. Other selectable markers, e.g., the

neomycin phosphotransferase gene, provide for isolation and identification of cells containing them by conferring properties on the cells that make them resistant to an agent, *e.g.*, a drug such as an antibiotic, that inhibits proliferation of cells that do not contain the marker.

- 5 As another example, interference of gene expression by double stranded RNA has been shown in *Caenorhabditis elegans*, plants, *Drosophila*, protozoans and mammals. This method is known as RNA interference (RNAi) and utilizes short, double-stranded RNA molecules (siRNAs). The siRNAs are generally composed of a 19-22bp double-
- 10 stranded RNA stem, a loop region and a 1-4 bp overhang on the 3' end. The reduction of gene expression has been accomplished by direct introduction of the siRNAs into the cell (Harborth J et al., 2001, J Cell Sci 114(pt 24):4557-65) as well as the introduction of DNA encoding and expressing the siRNA molecule. The encoded siRNA molecules are under
- 15 the regulation of an RNA polymerase III promoter (see, *e.g.*, Yu et al., 2002, Proc Natl Acad Sci USA 99(9):6047-52; Brummelkamp et al., 2002, Science 296(5567):550-3; Miyagishi et al., 2002, Nat Biotechnol 20(5):497-500; and the like). In certain embodiments, RNAi in mammalian cells may have advantages over other therapeutic methods.
- 20 For example, producing siRNA molecules that block viral genetic activities in infected cells may reduce the effects of the virus. Platform ACes provided herein encoding siRNA molecule(s) are an additional utilization of the platform ACes technology. The platform ACes could be engineered to encode one or more siRNA molecules to create gene "knockdowns". In
- 25 one embodiment, a platform ACes can be engineered to encode both the siRNA molecule and a replacement gene. For example, a mouse model or cell culture system could be generated using a platform ACes that has a knockdown of the endogenous mouse gene, by siRNA, and the human gene homolog expressing in place of the mouse gene. The placement of

siRNA encoding sequences under the regulation of a regulatable or inducible promoter would allow one to temporally and/or spatially control the knockdown effect of the corresponding gene.

As used herein, a reporter gene includes any gene that expresses a
5 detectable gene product, which may be RNA or protein. Generally
reporter genes are readily detectable. Examples of reporter genes include,
but are not limited to nucleic acid encoding a fluorescent protein, CAT
(chloramphenicol acetyl transferase) (Alton *et al.* (1979) *Nature* 282: 864-
869) luciferase, and other enzyme detection systems, such as beta-
10 galactosidase; firefly luciferase (deWet *et al.* (1987) *Mol. Cell. Biol.*
7:725-737); bacterial luciferase (Engebrecht and Silverman (1984) *Proc.*
Natl. Acad. Sci. U.S.A. 81:4154-4158; Baldwin *et al.* (1984)
Biochemistry 23:3663-3667); and alkaline phosphatase (Toh *et al.* (1989)
Eur. J. Biochem. 182:231-238, Hall *et al.* (1983) *J. Mol. Appl. Gen.*
15 2:101).

As used herein, growth under selective conditions means growth of a cell under conditions that require expression of a selectable marker for survival.

As used herein, an agent that destabilizes a chromosome is any
20 agent known by those skilled in the art to enhance amplification events,
and/or mutations. Such agents, which include BrdU, are well known to
those skilled in the art.

In order to generate an artificial chromosome containing a particular
heterologous nucleic acid of interest, it is possible to include the nucleic
25 acid in the nucleic acid that is being introduced into cells to initiate
production of the artificial chromosome. Thus, for example, a nucleic
acid can be introduced into a cell along with nucleic acid encoding a
selectable marker and/or a nucleic acid that targets to a heterochromatic
region of a chromosome. For introducing a heterologous nucleic acid into

the cell, it can be included in a fragment that includes a selectable marker or as part of a separate nucleic acid fragment and introduced into the cell with a selectable marker during the process of generating the artificial chromosomes. Alternatively, heterologous nucleic acid can be introduced
5 into an artificial chromosome at a later time after the initial generation of the artificial chromosome.

As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome that contains more euchromatic than heterochromatic DNA. For purposes herein, the
10 minichromosome contains a *de novo* centromere (e.g., a neocentromere). In some embodiments, for example, the minichromosome contains a centromere that replicates in animals, e.g., a mammalian centromere or in plants, e.g., a plant centromere.

As used herein, *in vitro* assembled artificial chromosomes or
15 synthetic chromosomes can be either more euchromatic than heterochromatic or more heterochromatic than euchromatic and are produced by joining essential components of a chromosome *in vitro*. These components include at least a centromere, a megareplicator, a telomere and optionally secondary origins of replication.

20 As used herein, *in vitro* assembled plant or animal artificial chromosomes are produced by joining essential components (at least the centromere, telomere(s), megareplicator and optional secondary origins of replication) that function in plants or animals. In particular embodiments, the megareplicator contains sequences of rDNA, particularly plant or
25 animal rDNA.

As used herein, a plant is a eukaryotic organism that contains, in addition to a nucleus and mitochondria, chloroplasts capable of carrying out photosynthesis. A plant can be unicellular or multicellular and can contain multiple tissues and/or organs. Plants can reproduce sexually or

asexually and can be perennial or annual in growth. Plants can also be terrestrial or aquatic. The term "plant" includes a whole plant, plant cell, plant protoplast, plant calli, plant seed, plant organ, plant tissue, and other parts of a whole plant.

5 As used herein, stable maintenance of chromosomes occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their
10 structure during cell culturing, suffering no unintended intrachromosomal or interchromosomal rearrangements.

 As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere in a chromosome as a result of incorporation of a heterologous nucleic acid fragment using the methods
15 herein.

 As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

20 As used herein, ribosomal RNA (rRNA) is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes (i.e., genes which encode rRNA)
25 per haploid genome are spread out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA (rDNA, which is DNA containing sequences that encode rRNA) has been verified on at least 11 pairs out of 20 mouse chromosomes (chromosomes 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, and 19)

(see e.g., Rowe *et al.* (1996) *Mamm. Genome* 7:886-889 and Johnson *et al.* (1993) *Mamm. Genome* 4:49-52). In *Arabidopsis thaliana* the presence of rDNA has been verified on chromosomes 2 and 4 (18S, 5.8S, and 25S rDNA) and on chromosomes 3,4, and 5 (5S rDNA)(see The
5 *Arabidopsis* Genome Initiative (2000) *Nature* 408:796-815). In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA
10 which can vary in length and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of
15 ribosome production within the cell nucleus.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of heterochromatin. Megachromosomes are made up of an array of repeated amplicons that contain two inverted megareplicons bordered by
20 introduced heterologous DNA (see, e.g., Figure 3 of U.S. Patent No. 6,077,697 for a schematic drawing of a megachromosome). For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes (about 90 to 120 or 150 Mb), dwarf
25 megachromosomes (~150-200 Mb), and a micro-megachromosome (~50-90 Mb, typically 50-60 Mb). For purposes herein, the term

megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments (amplicons) that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

5 As used herein, gene therapy involves the transfer or insertion of nucleic acid molecules into certain cells, which are also referred to as target cells, to produce specific products that are involved in preventing, curing, correcting, controlling or modulating diseases, disorders and deleterious conditions. The nucleic acid is introduced into the selected
10 target cells in a manner such that the nucleic acid is expressed and a product encoded thereby is produced. Alternatively, the nucleic acid may in some manner mediate expression of DNA that encodes a therapeutic product. This product may be a therapeutic compound, which is produced in therapeutically effective amounts or at a therapeutically
15 useful time. It may also encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Expression of the nucleic acid by the target cells within an organism afflicted with a disease or disorder thereby provides for modulation of the disease or disorder. The nucleic acid encoding the
20 therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

For use in gene therapy, cells can be transfected *in vitro*, followed by introduction of the transfected cells into an organism. This is often
25 referred to as *ex vivo* gene therapy. Alternatively, the cells can be transfected directly *in vivo* within an organism.

As used herein, therapeutic agents include, but are not limited to, growth factors, antibodies, cytokines, such as tumor necrosis factors and interleukins, and cytotoxic agents and other agents disclosed herein and

known to those of skill in the art. Such agents include, but are not limited to, tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), erythropoietin (EPO), pro-coagulants such as tissue factor and tissue factor variants, pro-apoptotic agents such as FAS-ligand, fibroblast growth factors (FGF), nerve growth factor and other growth factors.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures the disease.

As used herein, transgenic plants and animals refer to plants and animals in which heterologous or foreign nucleic acid is expressed or in which the expression of a gene naturally present in the plant or animal has been altered by virtue of introduction of heterologous or foreign nucleic acid.

As used herein, IRES (internal ribosome entry site; see, *e.g.*, SEQ ID No. 27 and nucleotides 2736-3308 SEQ ID No. 28) refers to a region of a nucleic acid molecule, such as an mRNA molecule, that allows internal ribosome entry sufficient to initiate translation, which initiation can be detected in an assay for cap-independent translation (see, *e.g.*, U.S. Patent No. 6,171,821). The presence of an IRES within an mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated.

Internal ribosome entry site (IRES) elements were first identified in picornaviruses, which elements are considered the paradigm for cap-independent translation. The 5' UTRs of all picornaviruses are long and mediate translational initiation by directly recruiting and binding
5 ribosomes, thereby circumventing the initial cap-binding step. IRES elements are frequently found in viral mRNA, they are rare in non-viral mRNA. Among non-viral mRNA molecules that contain functional IRES elements in their respective 5' UTRs are those encoding immunoglobulin heavy chain binding protein (BiP) (Macejak *et al.* (1991) *Nature*
10 353:90-94); *Drosophila* Antennapedia (Oh *et al.* (1992) *Genes Dev.* 6:1643-1653); *D. Ultrabithorax* (Ye *et al.* (1997) *Mol. Cell Biol.* 17:1714-21); fibroblast growth factor 2 (Vagner *et al.* (1995) *Mol. Cell Biol.* 15:35-44); initiation factor eIF4G (Gan *et al.* (1998) *J. Biol. Chem.* 273:5006-5012); proto-oncogene c-myc (Nanbru *et al.* (1995) *J. Biol.*
15 *Chem.* 272:32061-32066; Stoneley (1998) *Oncogene* 16:423-428); IRES_H; from the 5'UTR of NRF1 gene (Oumard *et al.* (2000) *Mol. and Cell Biol.*, 20(8):2755-2759); and vascular endothelial growth factor (VEGF) (Stein *et al.* (1998) *Mol. Cell Biol.* 18:3112-9).

As used herein, a promoter, with respect to a region of DNA, refers
20 to a sequence of DNA that contains a sequence of bases that signals RNA polymerase to associate with the DNA and initiate transcription of RNA (such as pol II for mRNA) from a template strand of the DNA. A promoter thus generally regulates transcription of DNA into mRNA. A particular promoter provided herein is the Ferritin heavy chain promoter (excluding
25 the Iron Response Element, located in the 5'UTR), which was joined to the 37bp Fer-1 enhancer element. This promoter is set forth as SEQ ID NO:128. The endogenous Fer-1 enhancer element is located upstream of the Fer-1 promoter (e.g., a Fer-1 oligo was cloned proximal to the core promoter).

As used herein, isolated, substantially pure nucleic acid, such as, for example, DNA, refers to nucleic acid fragments purified according to standard techniques employed by those skilled in the art, such as that found in Sambrook *et al.* ((2001) Molecular Cloning: A Laboratory
5 Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition).

As used herein, expression refers to the transcription and/or translation of nucleic acid. For example, expression can be the transcription of a gene that may be transcribed into an RNA molecule,
10 such as a messenger RNA (mRNA) molecule. Expression may further include translation of an RNA molecule and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA. With respect to an antisense
15 construct, expression may refer to the transcription of the antisense DNA.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous nucleic acids into cells for either expression of the heterologous nucleic acid or for replication of the heterologous nucleic acid. Selection and use of such vectors and
20 plasmids are well within the level of skill of the art.

As used herein, transformation/transfection refers to the process by which nucleic acid is introduced into cells. The terms transfection and transformation refer to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences
25 are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by *Agrobacterium*-mediated transformation, protoplast transformation (including polyethylene glycol (PEG)-mediated transformation, electroporation, protoplast fusion, and microcell fusion), lipid-mediated delivery, liposomes, electroporation,

- sonoporation, microinjection, particle bombardment and silicon carbide whisker-mediated transformation and combinations thereof (see, *e.g.*, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:169-177; Reich *et al.* (1986) *Biotechnology* 4:1001-1004; Klein *et al.* (1987) *Nature* 327:70-73; U.S. Patent No. 6,143,949; Paszkowski *et al.* (1989) in *Cell Culture and Somatic Cell Genetics of Plants, Vol. 6*, Molecular Biology of Plant Nuclear Genes, eds. Schell, J and Vasil, L.K. Academic Publishers, San Diego, California, p. 52-68; and Frame *et al.* (1994) *Plant J.* 6:941-948), direct uptake using
- 10 calcium phosphate (CaPO₄; see, *e.g.*, Wigler *et al.* (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76:1373-1376), polyethylene glycol (PEG)-mediated DNA uptake, lipofection (see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327), microcell fusion (see, EXAMPLES, see, also Lambert (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5907-5911; U.S. Patent No. 5,396,767,
- 15 Sawford *et al.* (1987) *Somatic Cell Mol. Genet.* 13:279-284; Dhar *et al.* (1984) *Somatic Cell Mol. Genet.* 10:547-559; and McNeill-Killary *et al.* (1995) *Meth. Enzymol.* 254:133-152), lipid-mediated carrier systems (see, *e.g.*, Teifel *et al.* (1995) *Biotechniques* 19:79-80; Albrecht *et al.* (1996) *Ann. Hematol.* 72:73-79; Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolch *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618) or other suitable method.
- 20 Methods for delivery of ACes are described in copending U.S. application Serial No. 09/815,979. Successful transfection is generally recognized
- 25 by detection of the presence of the heterologous nucleic acid within the transfected cell, such as, for example, any visualization of the heterologous nucleic acid or any indication of the operation of a vector within the host cell.

As used herein, "delivery," which is used interchangeably with "transfection," refers to the process by which exogenous nucleic acid molecules are transferred into a cell such that they are located inside the cell. Delivery of nucleic acids is a distinct process from expression of
5 nucleic acids.

As used herein, injected refers to the microinjection, such as by use of a small syringe, needle, or pipette, for injection of nucleic acid into a cell.

As used herein, substantially homologous DNA refers to DNA that
10 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids, with each other or a reference sequence, under specified conditions.

It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions,
15 hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 10, 14 or 16 or more nucleotides in a sequence that is complementary (or nearly complementary) to a substantially
20 contiguous sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with complementary base-pairing segments hybridize detectably to each other, departures from exact
25 complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids becomes larger, and as

conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence", if (a) both form a base-paired duplex with the same
5 segment, and (b) the melting temperatures of the two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of
10 nucleic acid duplexes are well known (see, *e.g.*, Meinkoth *et al.* (1984) *Anal. Biochem.* 138:267-284 and references cited therein).

As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes complementary or substantially complementary
15 sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt
20 concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

To be used as a hybridization probe, the nucleic acid is generally
25 rendered detectable by labeling it with a detectable moiety or label, such as ³²P, ³H and ¹⁴C, or by other means, including chemical labeling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

(via the biotin moieties) by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. (New York, NY). Any other label known to
5 those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available (for culturing cells, extracting DNA, and hybridization assays), the quantity of DNA or RNA available as a source of the probe, the particular label and the means used
10 to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques (see, *e.g.*, Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory Press).

15 As used herein, conditions under which DNA molecules form stable hybrids are considered substantially homologous, and a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a polypeptide. By the term "substantially homologous" is meant having at
20 least 75%, preferably 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be
25 determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970)), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)). Briefly, the GAP program defines similarity as the number of aligned symbols (*i.e.*,

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and
5 the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3)
10 no penalty for end gaps.

By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate
15 stringency or at high stringency all along the length of the nucleic acid of interest. Preferably the two molecules will hybridize under conditions of high stringency. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

20 Whether any two nucleic acid molecules have nucleotide sequences that are at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988).
25 Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine relative sequence identity.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

calculated using published techniques. (See, e.g.: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(11):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., et al., *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of

100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

10 As used herein: stringency of hybridization in determining percentage mismatch encompass the following conditions or equivalent conditions thereto:

- 1) high stringency: 0.1 x SSPE or SSC, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE or SSC, 0.1% SDS, 50°C
- 15 3) low stringency: 1.0 x SSPE or SSC, 0.1% SDS, 50°C

or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching. Equivalent conditions refer to conditions that select for substantially the same percentage of mismatch in the resulting hybrids. Additions of ingredients, such as formamide, Ficoll, and Denhardt's solution affect parameters such as the temperature under which the hybridization should be conducted and the rate of the reaction. Thus, hybridization in 5 X SSC, in 20% formamide at 42° C is substantially the same as the conditions recited above hybridization under conditions of low stringency. The recipes for SSPE, SSC and Denhardt's and the preparation of deionized formamide are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Chapter 8; see, Sambrook *et al.*, vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions. It is understood that

equivalent stringencies may be achieved using alternative buffers, salts and temperatures. As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

As used herein, conservative amino acid substitutions, such as those set forth in Table 1, are those that do not eliminate biological activity. Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
20	Arg (R)	Lys, orn
	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
25	Gly (G)	Ala; Pro
	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nva
	Lys (K)	Arg; Gln; Glu
30	Met (M)	Leu; Tyr; Ile; NLe Val
	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
35	Trp (W)	Tyr
	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nva

Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which
5 occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results
10 in more than one type of mRNA.

As used herein, a probe or primer based on a nucleotide sequence includes at least 10, 14, 16, 30 or 100 contiguous nucleotides from the reference nucleic acid molecule.

As used herein, recombinant production by using recombinant DNA
15 methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological
20 activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a
25 substrate, light is produced.

The terms substantially identical or similar varies with the context as understood by those skilled in the relevant art and generally means at least 40, 60, 80, 90, 95 or 98%.

As used herein, substantially identical to a product means
5 sufficiently similar so that the property is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel
10 electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce
15 substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector (or plasmid) refers to discrete elements that
20 are used to introduce heterologous DNA into cells for either expression or replication thereof. The vectors typically remain episomal, but may be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and
25 mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector

refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA.

Appropriate expression vectors are well known to those of skill in the art
5 and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein-binding-sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or
10 peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, a composition refers to any mixture of two or more ingredients. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

15 As used herein, a combination refers to any association between two or more items.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such
20 compositions.

As used herein, a cellular extract refers to a preparation or fraction that is made from a lysed or disrupted cell.

As used herein, the term "subject" refers to animals, plants, insects, and birds and other phyla, genera and species into which nucleic
25 acid molecules may be introduced. Included are higher organisms, such as mammals, fish, insects and birds, including humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, horses, chicken and others.

As used herein, flow cytometry refers to processes that use a laser based instrument capable of analyzing and sorting out cells and or chromosomes based on size and fluorescence.

As used herein, the abbreviations for any protective groups, amino
5 acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).

B. Recombination systems

10 Site-specific recombination systems typically contain three elements: a pair of DNA sequences (the site-specific recombination sequences) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction between two site-specific recombination sequences.

15 A number of different site-specific recombinase systems are available and/or known to those of skill in the art, including, but not limited to: the Cre/lox recombination system using CRE recombinase (see, *e.g.*, SEQ ID Nos. 58 and 59) from the Escherichia coli phage P1 (see, *e.g.*, Sauer (1993) *Methods in Enzymology* 225:890-900; Sauer *et al.* 20 (1990) *The New Biologist* 2:441-449), Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; Odell *et al.* (1990) *Mol Gen Genet.* 223:369-378; Lasko *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:6232-6236; U.S. Patent No. 5,658,772), the FLP/FRT system of yeast using the FLP recombinase (see, SEQ ID Nos. 60 and 61) from the 2 μ episome of 25 *Saccharomyces cerevisiae* (Cox (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:4223; Falco *et al.* (1982) *Cell* 29:573-584; Golic *et al.* (1989) *Cell* 59:499-509; U.S. Patent No. 5,744,336), the resolvases, including Gin recombinase of phage Mu (Maeser *et al.* (1991) *Mol Gen Genet.* 230:170-176; Klippel, A. *et al* (1993) *EMBO J.* 12:1047-1057; see, *e.g.*,

SEQ ID Nos. 64-67), Cin, Hin, $\alpha\delta$ Tn3; the Pin recombinase of *E. coli* (see, e.g., SEQ ID Nos. 68 and 69; Enomoto *et al.* (1983) *J. Bacteriol.* 6:663-668), the R/RS system of the pSR1 plasmid of *Zygosaccharomyces rouxii* (Araki *et al.* (1992) *J. Mol. Biol.* 225:25-37; Matsuzaki *et al.* (1990) *J. Bacteriol.* 172: 610-618) and site-specific recombinases from *Kluyveromyces drosophilarius* (Chen *et al.* (1986) *Nucleic Acids Res.* 314:4471-4481) and *Kluyveromyces waltii* (Chen *et al.* (1992) *J. Gen. Microbiol.* 138:337-345). Other systems are known to those of skill in the art (Stark *et al. Trends Genet.* 8:432-439; Utatsu *et al.* (1987) *J. Bacteriol.* 169:5537-5545; see, also, U.S. Patent No. 6,171,861).

Members of the highly related family of site-specific recombinases, the resolvase family, such as $\gamma\delta$, Tn3 resolvase, Hin, Gin, and Cin are also available. Members of this family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser *et al.* (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination (see, U.S. Patent No. 6,171,861).

The bacteriophage P1 Cre/lox and the yeast FLP/FRT systems are particularly useful systems for site-specific integration, inversion or excision of heterologous nucleic acid into, and out of, chromosomes, particularly ACes as provided herein. In these systems a recombinase (Cre or FLP) interacts specifically with its respective site-specific recombination sequence (lox or FRT, respectively) to invert or excise the intervening sequences. The sequence for each of these two systems is relatively short (34 bp for lox and 47 bp for FRT).

The FLP/FRT recombinase system has been demonstrated to function efficiently in plant cells (U.S. Patent No. 5,744,386), and, thus, can be used for producing plant artificial chromosome platforms. In

general, short incomplete FRT sites leads to higher accumulation of excision products than the complete full-length FRT sites. The system catalyzes intra- and intermolecular reactions, and, thus, can be used for DNA excision and integration reactions. The recombination reaction is
5 reversible and this reversibility can compromise the efficiency of the reaction in each direction. Altering the structure of the site-specific recombination sequences is one approach to remedying this situation. The site-specific recombination sequence can be mutated in a manner that the product of the recombination reaction is no longer recognized as
10 a substrate for the reverse reaction, thereby stabilizing the integration or excision event.

In the Cre-lox system, discovered in bacteriophage P1, recombination between loxP sites occurs in the presence of the Cre recombinase (see, *e.g.*, U.S. Patent No. 5,658,772). This system can be
15 used to insert, invert or excise nucleic acid located between two lox sites. Cre can be expressed from a vector. Since the lox site is an asymmetrical nucleotide sequence, lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA
20 segment located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule
25 result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the Cre coding region.

Any site-specific recombinase system known to those of skill in the art is contemplated for use herein. It is contemplated that one or a plurality of sites that direct the recombination by the recombinase are introduced into an artificial chromosome to produce platform ACes. The resulting platform ACes are introduced into cells with nucleic acid encoding the cognate recombinase, typically on a vector, and nucleic acid encoding heterologous nucleic acid of interest linked to the appropriate recombination site for insertion into the platform ACes. The recombinase-encoding-nucleic acid may be introduced into the cells on the same vector, or a different vector, encoding the heterologous nucleic acid.

An *E. coli* phage lambda integrase system for ACes platform engineering and for artificial chromosome engineering is provided (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181). The phage lambda integrase (Landy, A. (1989) *Annu. Rev. Biochem.* 58:913-94) is adapted herein and the cognate *att* sites are provided. Chromosomes, including ACes, engineered to contain one or a plurality of *att* sites are provided, as are vectors encoding a mutant integrase that functions in the absence other factors. Methods using the modified chromosomes and vectors for introduction of heterologous nucleic acid are also provided.

For purposes herein, one or more of the sites (e.g., a single site or a pair of sites) required for recombination are introduced into an artificial chromosome, such as an ACes chromosome. The enzyme for catalyzing site-directed recombination is introduced with the DNA of interest, or separately, or is engineered onto the artificial chromosome under the control of a regulatable promoter.

As described herein, artificial chromosome platforms containing one or multiple recombination sites are provided. The methods and resulting products are exemplified with the lambda phage Att/Int system, but

similar methods may be used for production of *ACes* platforms with other recombination systems.

The Att/Int system and vectors provided herein are not only intended for engineering *ACes* platforms, but may be used to engineer an Att/Int system into any chromosome. Introduction of att sites into a chromosome will permit engineering of natural chromosomes, such as by permitting targeted integration genes or regulatory regions, and by controlled excision of selected regions. For example, genes encoding a particular trait may be added to a chromosome, such as plant chromosome engineered to contain one or plurality of att sites. Such chromosomes may be used for screening DNA to identify genes. Large pieces of DNA can be introduced into cells and the cells screened phenotypically to select those having the desired trait.

C. Platforms

Provided herein are platform artificial chromosomes (platform *ACes*) containing single or multiple site-specific recombination sites. Chromosome-based platform technology permits efficient and tractable engineering and subsequent expression of multiple gene targets. Methods are provided that use DNA vectors and fragments to create platform artificial chromosomes, including animal, particularly mammalian, artificial chromosomes, and plant artificial chromosomes. The artificial chromosomes contain either single or multiple sequence-specific recombination sites suitable for the placement of target gene expression vectors onto the platform chromosome. The engineered chromosome-based platform *ACes* technology is applicable for methods, including cellular and transgenic protein production, transgenic plant and animal production and gene therapy. The platform *ACes* are also useful for producing a library of *ACes* comprising random portions of a given genome (e.g., a mammalian, plant or prokaryotic genome) for genomic

screening; as well as a library of cells comprising different and/or mutually exclusive *ACes* therein.

Exemplary of artificial chromosome platforms are those based on *ACes*. *ACes* artificial chromosomes are non-viral, self-replicating nucleic acid molecules that function as a natural chromosome, having all the elements required for normal chromosomal replication and maintenance within the cell nucleus. *ACes* artificial chromosomes do not rely on integration into the genome of the cell to be effective, and they are not limited by DNA carrying capacity and as such the therapeutic gene(s) of interest, including regulatory sequences, can be engineered into the *ACes*. In addition, *ACes* are stable *in vitro* and *in vivo* and can provide predictable long-term gene expression. Once engineered and delivered to the appropriate cell or embryo, *ACes* work independently alongside host chromosomes, for *ACes* that are predominantly heterochromatin producing only the products (proteins) from the genes it carries. As provided herein *ACes* are modified by introduction of recombination site(s) to provide a platform for ready introduction of heterologous nucleic acid. The *ACes* platforms can be used for production of transgenic animals and plants; as vectors for genetic therapy; for use as protein production systems; for animal models to identify and target new therapeutics; in cell culture for the development and production of therapeutic proteins; and for a variety of other applications.

1. Generation of artificial chromosomes

Artificial chromosomes may be generated by any method known to those of skill in the art. Of particular interest herein are the *ACes* artificial chromosomes, which contain a repeated unit. Methods for production of *ACes* are described in detail in U.S. Patent Nos. 6,025,155 and 6,077,697, which, as with all patents, applications, publications and other disclosure, are incorporated herein in their entirety.

Generation of *de novo* ACes.

ACes can be generated by cotransfecting exogenous DNA—such as a mammary tissue specific DNA cassette including the gene sequences for a therapeutic protein, with a rDNA fragment and a drug resistance
5 marker gene into the desired eukaryotic cell, such as plant or animal cells, such as murine cells *in vitro*. DNA with a selectable or detectable marker is introduced, and can be allowed to integrate randomly into pericentric heterochromatin or can be targeted to pericentric heterochromatin, such as that in rDNA gene arrays that reside on acrocentric chromosomes,
10 such as the short arms of acrocentric chromosomes. This integration event activates the “megareplicator” sequence and amplifies the pericentric heterochromatin and the exogenous DNA, and duplicates a centromere. Ensuing breakage of this “dicentric” chromosome can result in the production of daughter cells that contain the substantially-original
15 chromosome and the new artificial chromosome. The resulting ACes contain all the essential elements needed for stability and replication in dividing cells—centromere, origins of replications, and telomeres. ACes have been produced that express marker genes (lacZ, green fluorescent protein, neomycin-resistance, puromycin-resistance, hygromycin-
20 resistance) and genes of interest. Isolated ACes, for example, have been successfully transferred intact to rodent, human, and bovine cells by electroporation, sonoporation, microinjection, and transfection with lipids and dendrimers.

To render the creation of ACes with desired genes more tractable
25 and efficient, “platform” ACes (platform-ACes) can be produced that contain defined DNA sequences for enzyme-mediated homologous DNA recombination, such as by Cre or FLP recombinases (Bouhassira *et al.* (1996) *Blood* 88(supplement 1):190a; Bouhassira *et al.* (1997) *Blood*, 90:3332-3344; Siebler *et al.* (1997) *Biochemistry*: 36:1740-1747;

Siebler *et al.* (1998) *Biochemistry* 37: 6229-6234; and Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834), and as exemplified herein the lambda phage integrase. A *lox* site contains two 13 bp inverted repeats to which Cre-recombinase binds and an intervening 8 bp core region.

- 5 Only pairs of sites having identity in the central 6 bp of the core region are proficient for recombination; sites having non-identical core sequences (heterospecific *lox* sites) do not efficiently recombine with each other (Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300).

10 **Generating acrocentric chromosomes for plant artificial chromosome formation.**

In human and mouse cells *de novo* formation of a satellite DNA based artificial chromosome (SATAC, also referred to as ACes) can occur in an acrocentric chromosome where the short arm contains only pericentric heterochromatin, the rDNA array, and telomere sequences.

- 15 Plant species may not have any acrocentric chromosomes with the same physical structure described, but "megareplicator" DNA sequences reside in the plant rDNA arrays, also known as the nucleolar organizing regions (NOR). A structure like those seen in acrocentric mammalian chromosomes can be generated using site-specific recombination between
20 appropriate arms of plant chromosomes.

Approach

- Qin *et al.* ((1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:1706-1710, 1994) describes crossing two *Nicotiana tabacum* transgenic plants. One plant contains a construct encoding a promoterless hygromycin-resistance
25 gene preceded by a *lox* site (*lox-hpt*), the other plant carries a construct containing a cauliflower mosaic virus 35S promoter linked to a *lox* sequence and the *cre* DNA recombinase coding region (35S-*lox-cre*). The constructs were introduced separately by infecting leaf explants with *agrobacterium tumefaciens* which carries the kanamycin-resistance gene

(Kan^R). The resultant Kan^R transgenic plants were crossed. Plants that carried the appropriate DNA recombination event were identified by hygromycin-resistance.

5 **Modification of the above for generation of ACes**

 The Kan^R cultivars are initially screened, such as by FISH, to identify two sets of candidate transgenic plants. One set has one construct integrated in regions adjacent to the pericentric heterochromatin on the short arm of any chromosome. The second set of candidate plants
10 has the other construct integrated in the NOR region of appropriate chromosomes. To obtain reciprocal translocation both sites must be in the same orientation. Therefore a series of crosses are required, Kan^R plants generated, and FISH analyses performed to identify the appropriate "acrocentric" plant chromosome for *de novo* plant ACes formation.

15 **2. Bacteriophage lambda integrase-based site-specific recombination system**

 An integral part of the platform technology includes a site-specific recombination system that allows the placement of selected gene targets or genomic fragments onto the platform chromosomes. Any such system
20 may be used. In particular, a method is provided for insertion of additional DNA fragments into the platform chromosome residing in the cell via sequence-specific recombination using the recombinase activity of the bacteriophage lambda integrase. The lambda integrase system is exemplary of the recombination systems contemplated for ACes. Any
25 known recombination system, including any described herein, particularly any that operates without the need for additional factors or that, by virtue of mutation, does not require additional factors, is contemplated.

As noted the lambda integrase system provided herein can be used with natural chromosomes and artificial chromosomes in addition to ACes. Single or a plurality of recombination sites, which may be the same or different, are introduced into artificial chromosomes to produce
5 artificial chromosome platforms.

3. Creation of bacteriophage lambda integrase site-specific recombination system

The lambda phage-encoded integrase (designated Int) is a prototypical member of the integrase family. Int effects integration and
10 excision of the phage in and out of the *E. coli* genome via recombination between pairs of attachment sites designated *attB/attP* and *attL/attR*. Each *att* site contains two inverted 9 base pair core Int binding sites and a 7 base pair overlap region that is identical in wild-type *att* sites. Each
15 site, except for *attB* contains additional Int binding sites. In flanking regions, there are recognition sequences for accessory DNA binding proteins, such as integration host factor (IHF), factor for inversion stimulation (FIS) and the phage encoded excision protein (XIS). Except
20 for *attB*, Int is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins and negative DNA supercoiling, binds simultaneously to core and arm sites within the same *att* site.

Int, like Cre and FLP, executes an ordered sequential pair of strand exchanges during integrative and excisive recombination. The natural pairs of target sequences for Int, *attB* and *attP* or *attL* and *attR* are
25 located on the same or different DNA molecules resulting in intra or intermolecular recombination, respectively. For example, intramolecular recombination occurs between inversely oriented *attB* and *attP*, or between *attL* and *attR* sequences, respectively, leading to inversion of the intervening DNA segment.

Like the recombinase systems, such as Cre and FLP, Int directs site-specific recombination. Unlike the other systems, such as Cre and FLP, Int generally requires additional protein factors for integrative and excisive recombination and negative supercoiling for integrative recombination.

5 Hence, the Int system had not been used in eukaryotic targeting systems.

Mutant Int proteins, designated Int-h (E174K) and a derivative thereof Int-h/218(E174K/E218K) do not require accessory proteins to perform intramolecular integrative and excisive recombination in co-transfection assays in human cells (Lorbach *et al.* (2000) *J Mol. Biol.*
10 296:1175-1181); wild-type Int does not catalyze intramolecular recombination in human cells harboring target sites *attB* and *attP*. Hence it had been demonstrated that mutant Int can catalyze factor-independent recombination events in human cells.

There has been no demonstration by others that this system can be
15 used for engineering of eukaryotic genomes or chromosomes. Provided herein are chromosomes, including artificial chromosomes, such as but not limited to ACes that contain *att* sites (e.g., platform ACes), and the use of such chromosomes for targeted integration of heterologous DNA into such chromosomes in eukaryotic cells, including animal, such as
20 rodent and human, and plant cells. Mutant Int provided herein is shown to effect site-directed recombination between sites in artificial chromosomes and vectors containing cognate sites.

An additional component of the chromosome-based platform technology is the site-specific integration of target DNA sequences onto
25 the platform. For this the native bacteriophage lambda integrase has been modified to carry out this sequence specific DNA recombination event in eukaryotic cells. The bacteriophage lambda integrase and its cognate DNA substrate *att* is a member of the site-specific recombinase family that also includes the bacteriophage P1 Cre/lox system as well as

the *Saccharomyces cerevisiae* 2 micron based FLP/FRT system (see, e.g., Landy (1989) *Ann. Rev. Biochem* 58:913-949; Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402; Broach *et al.* (1982) *Cell* 29:227-234).

5 By combining DNA endonuclease and DNA ligase activity these recombinases recognize and catalyze DNA exchanges between sequences flanking the recognition site. During the integration of lambda genome into the *E. coli* (lambda recombination) genome, the phage integrase (INT) in association with accessory proteins catalyzes the DNA exchange
10 between the attP site of the phage genome and the attB site of the bacterial genome resulting in the formation of attL and attR sites (Figure 6). The engineered bacteriophage lambda integrase has been produced herein to carry out an intermolecular DNA recombination event between
15 an incoming DNA molecule (primarily on a vector containing the bacterial attB site) and the chromosome-based platform carrying the lambda attP sequence independent of lambda bacteriophage or bacterial accessory proteins.

 In contrast to the bi-directional Cre/lox and FLP/FRT system, the engineered lambda recombination system derived for chromosome-based
20 platform technology is advantageously unidirectional because accessory proteins, which are absent, are required for excision of integrated nucleic acid upon further exposure to the lambda Int recombinase.

4. Creation of platform chromosome containing single or multiple sequence-specific recombination sites

25 **a. Multiple sites**

 For the creation of a platform chromosome containing multiple, sequence-specific recombination sites, artificial chromosomes are produced as depicted in Figure 5 and Example 3. As discussed above, artificial chromosomes can be produced using any suitable methodology,

including those described in U.S. Patent Nos. 5,288,625; 5,712,134; 5,891,691; 6,025,155. Briefly, to prepare artificial chromosomes containing multiple recombination (e.g., integration) sites, nucleic acid (either in the form a one or more plasmids, such as the plasmid pSV40193attPsensePUR set forth in Example 3) is targeted into an amplifiable region of a chromosome, such as the pericentric region of a chromosome. Among such regions are the rDNA gene loci in acrocentric mammalian chromosomes. Hence, targeting nucleic acid for integration into the rDNA region of mammalian acrocentric chromosomes can include the mouse rDNA fragments (for targeting into rodent cell lines) or large human rDNA regions on BAC/PAC vectors (or subclones thereof in standard vectors) for targeting into human acrocentric chromosomes, such as for human gene therapy applications. The targeting nucleic acid generally includes a detectable or selectable marker, such as antibiotic resistance, such as puromycin and hygromycin, a recombination site (such as attP, attB, attL, attR or the like), and/or human selectable markers as required for gene therapy applications. Cells are grown under conditions that result in amplification and ultimately production of ACes artificial chromosomes having multiple recombination (e.g., integration) sites therein. ACes having the desired size are selected for further engineering.

b. Creation of platform chromosome containing a single sequence-specific recombination site

In this method a mammalian platform artificial chromosome is generated containing a single sequence-specific recombination site. In the Example below, this approach is demonstrated using a puromycin resistance marker for selection and a mouse rDNA fragment for targeting into the rDNA locus on mouse acrocentric chromosomes. Other selection markers and targeting DNA sequences as desired and known to those of

skill in the art can be used. Additional resistance markers include genes conferring resistance to the antibiotics neomycin, blasticidin, hygromycin and zeocin. For applications, such as gene therapy in which potentially immunogenic responses are to be avoided, host, such as human, derived

5 selectable markers or markers detectable with monoclonal antibodies (MAb) followed by fluorescent activated cell sorting (FACS) can be used. Examples in this class include, but are not limited to: human nerve growth factor receptor (detection with MAb); truncated human growth factor receptor (detection with MAb); mutant human dihydrofolate reductase

10 (DHFR; detectable using a fluorescent methotrexate substrate); secreted alkaline phosphatase (SEAP; detectable with fluorescent substrate); thymidylate synthase (TS; confers resistance to fluorodeoxyuridine); human CAD gene (confers resistance to N-phosphonacetyl-L-aspartate (PALA)).

15 To construct a platform artificial chromosome with a single site, an ACes artificial chromosome (or other artificial chromosome of interest) can be produced containing a selectable marker. A single sequence specific recombination site is targeted onto ACes via homologous recombination. For this, DNA sequences containing the site-specific

20 recombination sequence are flanked with DNA sequences homologous to a selected sequence in the chromosome. For example, when using a chromosome containing rDNA or satellite DNA, such DNA can be used as homologous sequences to target the site-specific recombination sequence onto the chromosome. A vector is designed to have these homologous

25 sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the chromosome, the DNA is transfected into cells harboring the chromosome. After transfection and integration of the site-specific cassette, homologous recombination events onto the platform

chromosome are subcloned and identified, for example by screening single cell subclones via expression of resistance or a fluorescent marker and PCR analysis. In one embodiment, a platform artificial chromosome, such as a platform ACes, that contains a single copy of the recombination site is selected. Examples 2B and 2D exemplify the process, and Figure 3 provides a diagram depicting one method for the creation of a platform mammalian chromosome containing a single sequence-specific recombination site.

5 **5. Lambda integrase mediated recombination of target gene expression vector onto platform chromosome**

10 The third component of the chromosome-based platform technology involves the use of target gene expression vectors carrying, for example, genes for gene therapy, genes for transgenic animal or plant production, and those required for cellular protein production of interest.

15 Using lambda integrase mediated site-specific recombination, or any other recombinase-mediated site-specific recombination, the target gene expression vectors are introduced onto the selected chromosome platform. The use of target gene expression vector permits use of the *de novo* generated chromosome-based platforms for a wide range of gene

20 targets. Furthermore, chromosome platforms containing multiple *attP* sites provides the opportunity to incorporate multiple gene targets onto a single platform, thereby providing for expression of multiple gene targets, including the expression of cellular and genetic regulatory genes and the expression of all or parts of metabolic pathways. In addition to

25 expressing small target genes, such as cDNA and hybrid cDNA/artificial intron constructs, the chromosome-based platform can be used for engineering and expressing large genomic fragments carrying target genes along with its endogenous genomic promoter sequences. This is of importance, for example, where the therapy requires precise cell specific

expression and in instances where expression is best achieved from genomic clones rather than cDNA clones. Figure 9 provides a diagram summarizing one embodiment of the chromosome-based technology.

A feature of the target gene expression vector that is of interest to include is a promoterless marker gene, which as exemplified (see, Figure 9) contains an upstream *attB* site (marker 2 on Figure 9). The nucleic acid encoding the marker is not expressed unless it is placed downstream from a promoter sequence. Using the recombinase technology provided herein, such as the lambda integrase technology (λ INT_{E174R} on figure 8) provided herein, site-specific recombination between the *attB* site on the vector and the promoter-*attP* site (in the "sense" orientation) on the chromosome-based platform results in the expression of marker 2 on the target gene expression vector, thereby providing a positive selection for the lambda INT mediated site-specific recombination event. Site-specific recombination events on the chromosome-based platform versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. Examples of suitable marker 2 genes, include, but are not limited to, genes that confer resistance to toxic compounds or antibiotics, fluorescence activated cell sorting (FACS) sortable cell surface markers and various fluorescent markers. Examples of these genes include, but are not limited to, human L26a^R (human homolog of *Saccharomyces cerevisiae* CYH^B gene), neomycin, puromycin, blasticidin, CD24 (see, e.g., US Patents 5,804,177 and 6,074,836), truncated CD4, truncated low affinity nerve growth factor receptor (LNGFR), truncated LDL receptor, truncated human growth hormone receptor, GFP, RFP, BFP.

The target gene expression vectors contain a gene (target gene) for expression from the chromosome platform. The target gene can be expressed using various constitutive or regulated promoter systems

across various mammalian species. For the expression of multiple target genes within the same target gene expression vector, the expression of the multiple targets can be coordinately regulated via viral-based or human internal ribosome entry site (IRES) elements (see, *e.g.*, Jackson *et al.* (1990) *Trends Biochem Sci.* 15: 477-83; Oumard *et al.* (2000) *Mol. Cell. Biol.* 20: 2755-2759). Furthermore, using IRES type elements linked to a downstream fluorescent marker, *e.g.*, green, red or blue fluorescent proteins (GFP, RFP, BFP) allows for the identification of high expressing clones from the integrated target gene expression vector.

- 10 In certain embodiments described herein, the promoterless marker can be transcriptionally downstream of the heterologous nucleic acid, wherein the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.
- 15 In addition, the selectable marker and the heterologous nucleic acid can be transcriptionally linked by the presence of a IRES between them. As set forth herein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.
- 20 Expression from the target gene expression vector integrated onto the chromosome-based platform can be further enhanced using genomic insulator/boundary elements. The incorporation of insulator sequences into the target gene expression vector helps define boundaries in chromatin structure and thus minimizes influence of chromatin position
- 25 effects/gene silencing on the expression of the target gene (Bell *et al.* (1999) *Current Opinion in Genetics and Development* 9:191-198; Emery *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:9150-9155). Examples of insulator elements that can be included onto target gene expression vector in order to optimize expression include, but are not limited to:

- 1) chicken β -globin HS4 element (Prioleau *et al.* (1999) *EMBO J* 18: 4035-4048);
- 2) matrix attachment regions (MAR; see, *e.g.*, Ramakrishnan *et al.* (2000) *Mol Cell. Biol.* 20:868-877);
- 5 3) scaffold attachment regions (SAR; see, *e.g.*, Auten *et al.* (1999) *Human Gene Therapy* 10:1389-1399); and
- 4) universal chromatin opening elements (UCOE; WO/0005393 and WO/0224930)

The copy number of the target gene can be controlled by sequentially adding multiple target gene expression vectors containing the target gene onto multiple integration sites on the chromosome platform. Likewise, the copy number of the target gene can be controlled within an individual target gene expression vector by the addition of DNA sequences that promote gene amplification. For example, gene amplification can be induced utilizing the dihydrofolate reductase (DHFR) minigene with subsequent selection with methotrexate (see, *e.g.*, Schimke (1984) *Cell* 37:705-713) or amplification promoting sequences from the rDNA locus (see, *e.g.*, Wegner *et al.* (1989) *Nucl. Acids Res.* 17: 9909-9932).

20 6. Platforms with other recombinase system sites

A "double *lox*" targeting strategy mediated by Cre-recombinase (Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834) can be used. This strategy employs a pair of heterospecific *lox* sites—*loxA* and *loxB*, which differ by one nucleotide in the 8 bp spacer region. Both sites are engineered into the artificial chromosome and also onto the targeting DNA vector. This allows for a direct site-specific insertion of a commercially relevant gene or genes by a Cre-catalyzed double crossover event. In essence a platform ACes is engineered with a hygromycin-resistance gene flanked by the double *lox* sites generating *lox-ACes*, which is maintained

in the thymidine kinase deficient cell, LMtk(-). The gene of interest, for example, for testing purposes, the green fluorescence protein gene, GFP and a HSV thymidine kinase gene (tk) marker, are engineered between the appropriate *lox* sites of the targeting vector. The vector DNA is
5 cotransfected with plasmid pBS185 (Life Technologies) encoding the *Cre* recombinase gene into mammalian cells maintaining the dual-*lox* artificial chromosome. Transient expression of the *Cre* recombinase catalyzes the site-specific insertion of the gene and the tk-gene onto the artificial chromosome. The transfected cells are grown in HAT medium that
10 selects for only those cells that have integrated and expressed the thymidine kinase gene. The HAT^R colonies are screened by PCR analyses to identify artificial chromosomes with the desired insertion.

To generate the *lox-ACes*, Lambda-Hyg^R-*lox* DNA is transfected into the LMtk(-) cell line harboring the precursor *ACes*. Hygromycin-
15 resistant colonies are analyzed by FISH and Southern blotting for the presence of a single copy insert on the *ACes*.

To demonstrate the gene replacement technology, cell lines containing candidate *lox-ACes* are cotransfected with pTK-GFP-*lox* and pBS185 (encoding the *Cre* recombinase gene) DNA. After transfection,
20 transient expression of plasmid pBS185 will provide sufficient burst of *Cre* recombinase activity to catalyze DNA recombination at the *lox* sites. Thus, a double crossover event between the *ACes* target and the exogenous targeting plasmid carrying the *loxA* and *loxB* permits the simple replacement of the hygromycin-resistance gene on the *lox-ACes*
25 for the tk-GFP cassette from the targeting plasmid, with no integration of vector DNA. Transfected cells are grown in HAT-media to select for tk-expression. Correct targeting will result in the generation of HAT^R, hygromycin sensitive, and green fluorescent cells. The desired integration event is verified by Southern and PCR analyses. Specific PCR primer sets

are used to amplify DNA sequences flanking the individual *loxA* and *loxB* sites on the *lox-ACes* before and after homologous recombination.

D. Exemplary applications of the Platform *ACes*

Platform *ACes* are applicable and tractable for different/optimized
5 cell lines. Those that include a fluorescent marker, for example, can be purified and isolated using fluorescent activated cell sorting (FACS), and subsequently delivered to a target cell. Those with selectable markers provide for efficient selection and provide a growth advantage. Platform *ACes* allow multiple payload delivery of donor target vectors via a
10 positive-selection site-specific, recombination system, and they allow for the inclusion of additional genetic factors that improve protein production and protein quality.

The construction and use of the platform *ACes* as provided for each application may be similarly applied to other applications. Particular
15 descriptions are for exemplification.

1. Cellular Protein Production Platform *ACes* (CPP *ACes*)

As described herein, *ACes* can be produced from acrocentric chromosomes in rodent (mouse, hamster) cell lines via megareplicator induced amplification of heterochromatin/rDNA sequences. Such *ACes*
20 are ideal for cellular protein production as well as other applications described herein and known to those of skill in the art. *ACes* platforms that contain a plurality of recombination sites are particularly suitable for engineering as cellular protein production systems.

In one embodiment, CPP *ACes* involve a two-component system:
25 the platform chromosome containing multiple engineering sites and the donor target vector containing a platform-specific recombination site with designed expression cassettes (see Figure 9).

The platform *ACes* can be produced from any artificial chromosome, particularly the amplification-based artificial chromosomes.

For exemplification, they are produced from rodent artificial chromosomes produced from acrocentric chromosomes using the technology of U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183, in which nucleic acid is targeted to the pericentric heterochromatic, and, particularly into rDNA to initiate the replication event(s). The ACes can be produced directly in the chosen cellular protein production cell lines, such as, but not limited to, CHO cells, hybridomas, plant cells, plant tissues, plant protoplasts, stem cells and plant calli.

10 **a. Platform Construction**

 In the exemplary embodiment, the initial *de novo* platform construction requires co-transfecting with excess targeting DNA, such as, rDNA or lambda DNA without an *attP* region, and an engineered selectable marker. The engineered selectable marker should contain
15 promoter, generally a constitutive promoter, such as human, viral, i.e., adenovirus or SV40 promoter, including the human ferritin heavy chain promoter (SEQ ID NO:128), SV40 and EF1 α promoters, to control expression of a marker gene that provides a selective growth advantage to the cell. An example of such a marker gene is the *E. coli hisD* gene
20 (encoding histidinol dehydrogenase) which is homologous and analogous to the *S. typhimurium hisD* a dominant marker selection system for mammalian cells previously described (see, Hartman *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051). Since histidine is an essential amino acid in mammals and a nutritional requirement in cell culture, the *E.*
25 *coli hisD* gene can be used to select for histidine prototrophy in defined media. Furthermore more stringent selection can be placed on the cells by including histinol in the medium. Histidinol is itself permeable and toxic to cells. The *hisD* provides a means of detoxification.

Placed between the promoter and the marker gene is the bacteriophage lambda *attP* site to use the bacteriophage lambda integrase dependent site-specific recombination system (described herein). The insertion of an *attP* site downstream of a promoter element provide
5 forward selection of site-specific recombination events onto the platform *ACes*.

b. Donor Target Vector Construction

A second component of the CPP platform *ACes* system involves the construction of donor target vectors containing a gene product(s) of
10 interest for the CPP platform *ACes*. Individual donor target vectors can be designed for each gene product to be expressed thus enabling maximum usage of a *de novo* constructed platform *ACes*, so that one or a few CPP platform *ACes* will be required for many gene targets.

A key feature of the donor vector target is the *promoterless* marker
15 gene containing an upstream *attB* site (marker 2 on figure 9). Normally the marker would not be expressed unless it is placed downstream of a promoter sequence. As discussed above, using the lambda integrase technology (λ INT_{E174R} on Figure 8 and Figure 9), site-specific recombination between the *attB* site on the vector and the promoter-*attP*
20 site on the CPP platform *ACes* result in the expression of the donor target vector marker providing positive selection for the site-specific event. Site-specific recombination events on the CPP *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.
25 In addition, since the lambda integrase reaction is unidirectional, i.e. excision reaction is not possible, a number of unique targets can be loaded onto the CPP platform *ACes* limited only by the number of markers available.

Additional features of the donor target vector include gene target expression cassettes flanked by either chromatin insulator regions, matrix attachment regions (MAR) or scaffold attachment regions (SAR). The use of these regions will provide a more "open" chromatin environment for gene expression and help alleviate silencing. An example of such a cassette for expressing a monoclonal antibody is described. For this purpose, a strong constitutive promoter, e.g. chicken β -actin or RNA Poll, is used to drive the expression of the heavy and light chain open reading frames. The heavy and light chain sequences flank a nonattenuated human IRES (IRES_H; from the 5'UTR of NRF1 gene; see Oumard et al., 2000, *Mol. and Cell Biol.*, 20(8):2755-2759) element thereby coordinating transcription of both heavy and light chain sequence. Distal to the light chain open reading frame resides an additional viral encoded IRES (IRES_V modified ECMV internal ribosomal entry site (IRES)) element attenuating the expression of the fluorescent marker gene hrGFP from *Renilla* (Stratagene). By linking the hrGFP with an attenuated IRES, the heavy and light chains along with the hrGFP are monocistronic. Thus, the identification of hrGFP fluorescing cells will provide a means to detect protein producing cells. In addition, high producing cell lines can be identified and isolated by FACS thereby decreasing the time frame in finding high expressers. Functional monoclonal antibody will be confirmed by ELISA.

c. Additional components in cellular protein production platform ACes (CPP ACes)

In addition to the aforementioned CPP ACes system, other genetic factors can be included to enhance the yield and quality of the expressed protein. Again to provide maximum flexibility, these additional factors can be inserted onto the CPP platform ACes by λ INTE174R dependent site-specific recombination. Other factors that could be used with a CPP

Platform ACes include for example, adenovirus E1a transactivation system which upregulates both cellular and viral promoters (see, e.g., Svensson and Akusjarvi (1984) EMBO 3:789-794; and US patents 5,866,359; 4,775,630 and 4,920,211).

5 d. **Targets for CHO-ACes engineering to enhance cell growth, such as CHO cell growth and protein production/ quality**

 If adding these additional factors onto the CPP ACes is not prudent or desired, the host cell, CHO cells, can be engineered to express these
10 factors (see, below, targets for CHO-ACes engineering to enhance CHO cell growth and protein production/quality). Additional factors to consider including are addition of insulin or IGF-1 to sustain viability; human sialyltransferases or related factors to produce more human-like glycoproteins; expression of factors to decrease ammonium accumulation
15 during cell growth; expression of factors to inhibit apoptosis; expression of factors to improve protein secretion and protein folding; and expression of factors to permit serum-free transfection and selection.

 1) **Addition of insulin or IGF-1 to sustain viability**

20 Stimulatory factors and/or their receptors are expressed to set up an autocrine loop, to improve cell growth, such as CHO cell growth. Two exemplary candidates are insulin and IGF-1 (see, Biotechnol Prog 2000 Sep;16(5):693-7). Insulin is the most commonly used growth factor for sustaining cell growth and viability in serum-free Chinese hamster ovary
25 (CHO) cell cultures. Insulin and IGF-1 analog (LongR(3) serve as growth and viability factors for CHO cells.

 CHO cells were modified to produce higher levels of essential nutrients and factors. A serum-free (SF) medium for dihydrofolate reductase-deficient Chinese hamster ovary cells (DG44 cells) was
30 prepared. Chinese hamster ovary cells (DG44 cells), which are normally

maintained in 10% serum medium, were gradually weaned to 0.5% serum medium to increase the probability of successful growth in SF medium (see, Kim *et al.* (199) *In Vitro Cell Dev Biol Anim* 35(4):178-82). A SF medium (SF-DG44) was formulated by supplementing the basal medium with these components; basal medium was prepared by supplementing Dulbecco's modified Eagle's medium and Ham's nutrient mixture F12 with hypoxanthine (10 mg/l) and thymidine (10 mg/l). Development of a SF medium for DG44 cells was facilitated using a Plackett-Burman design technique and weaning of cells.

10

2) Human sialyltransferases or related factors to produce more human-like glycoproteins

CHO cells have been modified by increasing their ability to process protein via addition of complex carbohydrates. This has been achieved by overexpression of relevant processing enzymes, or in some cases, reducing expression of relevant enzymes (see, Bragonzi *et al.* (2000) *Biochim Biophys Acta* 1474(3):273-282; see, also Weikert *et al.* (1999) *Nature biotech.* 17:1116-11121; Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). A CHO cell line expressing alpha2,6-sialyltransferase was developed for the production of human-like sialylated recombinant glycoproteins. The sialylation defect of CHO cells can be corrected by transfecting the alpha2,6-sialyltransferase (alpha2,6-ST) cDNA into the cells. Glycoproteins produced by such CHO cells display alpha2,6-and alpha2,3-linked terminal sialic acid residues, similar to human glycoproteins.

As another example for improving the production of human-like sialylated recombinant glycoproteins, a CHO cell line has been developed that constitutively expresses sialidase antisense RNA (see, Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). Several antisense expression

30

vectors were prepared using different regions of the sialidase gene. Co-transfection of the antisense constructs with a vector conferring puromycin resistance gave rise to over 40 puromycin resistant clones that were screened for sialidase activity. A 5' 474 bp coding segment of the sialidase cDNA, in the inverted orientation in an SV 40-based expression vector, gave maximal reduction of the sialidase activity to about 40% wild-type values.

Oligosaccharide biosynthesis pathways in mammalian cells have been engineered for generation of recombinant glycoproteins (see, *e.g.*, Sburlati (1998) *Biotechnol Prog* 14(2):189-92), which describes a Chinese hamster ovary (CHO) cell line capable of producing bisected oligosaccharides on glycoproteins. This cell line was created by overexpression of a recombinant N-acetylglucosaminyltransferase III (GnT-III) (see, also, Prati *et al.* (1998) *Biotechnol Bioeng* 59(4):445-50, which describes antisense strategies for glycosylation engineering of CHO cells).

3) Expression of factors to decrease ammonium accumulation during cell growth

Excess ammonium, which is a by-product of CHO cell metabolism can have detrimental effects on cell growth and protein quality (see, Yang *et al.* (2000) *Biotechnol Bioeng* 68(4):370-80). To solve this problem ammonium levels were modified by overexpressing carbamoyl phosphate synthetase I and ornithine transcarbamoylase or glutamine synthetase in CHO cells. Such modification resulted in reduced ammonium levels observed and an increase in the growth rate (see Kim *et al.* (2000) *J Biotechnol* 81(2-3):129-40; and Enosawa *et al.* (1997) *Cell Transplant* 6(5):537-40).

4) Expression of factors to improve protein secretion and protein folding

Overexpression of relevant enzymes can be engineered into the ACes to improve protein secretion and folding.

5) Expression of factors to permit serum-free transfection and selection

5 It is advantageous to have the ability to convert CHO cells in suspension growing in serum free medium to adherence with out having to resort to serum addition. Laminin or fibronectin addition is sufficient to make cells adherent (see, *e.g.*, Zaworski *et al.* (1993) *Biotechniques* 15(5):863-6) so that expressing either of these genes in CHO cells under
10 an inducible promoter should allow for reversible shift to adherence without requiring serum addition.

2. Platform ACes and Gene Therapy

The platform ACes provided herein are contemplated for use in mammalian gene therapy, particularly human gene therapy. Human ACes
15 can be derived from human acrocentric chromosomes from human host cells, in which the amplified sequences are heterochromatic and/or human rDNA. Different platform ACes applicable for different tissue cell types are provided. The ACes for gene therapy can contain a single copy of a therapeutic gene inserted into a defined location on platform ACes.
20 Therapeutic genes include genomic clones, cDNA, hybrid genes and other combinations of sequences. Preferred selectable markers are those from the mammalian host, such as human derived factors so that they are non-immunogenic, non-toxic and allow for efficient selection, such as by FACS and/or drug resistance.

25 Platform ACes, useful for gene therapy and other applications, as noted herein, can be generated by megareplicator dependent amplification, such as by the methods in U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183. In one embodiment, human ACes are produced using

human rDNA constructs that target rDNA arrays on human acrocentric chromosomes and induce the megareplicator in human cells, particularly in primary cell lines (with sufficient number of doublings to form the ACes) or stem cells (such as hematopoietic stem cells, mesenchymal stem cells, adult stem cells or embryonic stem cells) to avoid the introduction of potentially harmful rearranged DNA sequences present in many transformed cell lines. Megareplicator induced ACes formation can result in multiple copies of targeting DNA/selectable markers in each amplification block on both chromosomal arms of the platform ACes.

10 In view of the considerations regarding immunogenicity and toxicity, the production of human platform ACes for gene therapy applications employs a two component system analogous to the platform ACes designed for cellular protein production (CPP platform ACes). The system includes a platform chromosome of entirely human DNA origin
15 containing multiple engineering sites and a gene target vector carrying the therapeutic gene of interest.

a. Platform Construction

The initial *de novo* construction of the platform chromosome employs the co-transfection of excess targeting DNA and a selectable
20 marker. In one embodiment, the DNA is targeted to the rDNA arrays on the human acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22). For example, two large human rDNA containing PAC clones 18714 and 18720 and the human PAC clone 558F8 are used for targeting (Genome Research (ML) now Incyte, BACPAC Resources, 747 52nd
25 Street, Oakland CA). The mouse rDNA clone pFK161 (SEQ ID NO: 118), which was used to make the human SATAC from the 94-3 hamster/human hybrid cell line (see, *e.g.*, published International PCT application No. WO 97/40183 and Csonka, *et al*, *Journal of Cell Science*

113:3207-32161 and Example 1 for a description of pFK161) can also be used.

For animal applications, selectable markers should be non-immunogenic in the animal, such as a human, and include, but are not limited to: human nerve growth factor receptor (detected with a MAb, such as described in US patent 6,365,373); truncated human growth factor receptor (detected with MAb), mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine); human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34+ cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS); human CD25 (IL-2 α ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); and Cytidine deaminase (CD; selectable by Ara-C).

Since megareplicator induced amplification generates multiple copies of the selectable marker, a second consideration for the selection of the human marker is the resulting dose of the expressed marker after ACes formation. High level of expression of certain markers may be detrimental to the cell and/or result in autoimmunity. One method to decrease the dose of the marker protein is by shortening its half-life, such as via the fusion of the well-conserved human ubiquitin tag (a 76 amino acid sequence) thus leading to increased turnover of the selectable marker. This has been used successfully for a number of reporter

systems including DHFR (see, *e.g.*, Stack *et al.* (2000) *Nature Biotechnology* 18:1298-1302 and references cited therein).

Using the ubiquitin tagged protein, a human selectable marker system analogous to the CPP ACes described herein is constructed.

- 5 Briefly, a tagged selectable marker, such as for example one of those described herein, is cloned downstream of an *attP* site and expressed from a human promoter. Exemplary promoters contemplated for use herein include, but are not limited to, the human ferritin heavy chain promoter (SEQ ID NO:128); RNA Pol; EF1 α ; TR; glyceraldehyde-3-
10 phosphate dehydrogenase core promoter (GAP); a GAP core promoter including a proximal insulin inducible element and the intervening GAP sequence; phosphofructokinase promoter; and phosphoglycerate kinase promoter. Also contemplated herein is an aldolase A promoter H1 & H2 (representing closely spaced transcriptional start sites) along with the
15 proximal H enhancer. There are 4 promoters (*e.g.*, transcriptional start sites) for this gene, each having different regulatory and tissue activity. The H (most proximal 2) promoters are ubiquitously expressed off the H enhancer. This resulting marker can then be co-transfected along with excess human rDNA targeting sequence into the host cells. An important
20 criteria for the selection of the

- recipient cells is sufficient number of cell doublings for the formation and detection of ACes. Accordingly, the co-transfections should be attempted in human primary cells that can be cultured for long periods of
25 time, such as for example, stem cells (*e.g.*, hematopoietic, mesenchymal, adult or embryonic stem cells), or the like. Additional cell types, include, but are not limited to: single gene transfected cells exhibiting increased life-span; over-expressing c-myc cells, *e.g.* MSU1.1 (Morgan *et al.*, 1991, *Exp. Cell Res.*, Nov;197(1):125-136); over-expressing telomerase lines,

such as TERT cells; SV40 large T-antigen transfected lines; tumor cell lines, such as HT1080; and hybrid human cell lines, such as the 94-3 hamster/human hybrid cell line.

b. Gene Target Vector

5 The second component of the GT platform *ACes* (GT *ACes*) system involves the use of engineered target vectors carrying the therapeutic gene of interest. These are introduced onto the GT platform *ACes* via site-specific recombination. As with the CPP *ACes*, the use of engineered target vectors maximizes the use of the *de novo* generated GT platform
10 *ACes* for most gene targets. Furthermore, using lambda integrase technology, GT platform *ACes* containing multiple *attP* sites permits the opportunity to incorporate multiple therapeutic targets onto a single platform. This could be of value in cases where a defined therapy requires multiple gene targets, a single therapeutic target requires an
15 additional gene regulatory factor or a GT *ACes* requires a "kill" switch.

 Similar to the CPP *ACes*, a feature of the gene target vector is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on Figure 9). Normally, the marker (in this case, a cell surface antigen that can be sorted by FACS would be ideal) would not be expressed unless it
20 is placed downstream of a promoter sequence. Using the lambda integrase technology (λ INT_{E174R} on figure 9), site-specific recombination between the *attB* site on the vector and the promoter- *attP* site on the GT platform *ACes* results in the expression of marker#2 on the gene target vector, i.e. positive selection for the site-specific event. Site-specific
25 recombination events on the GT *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.

 For expression of the therapeutic gene, human specific promoters, such as a ferritin heavy chain promoter (SEQ ID NO:128); EF1 α or RNA

Poll, are used. These promoters are for high level expression of a cDNA encoded therapeutic protein. In addition to expressing cDNA (or even hybrid cDNA/artificial intron constructs), the GT platform *ACes* are used for engineering and expressing large genomic fragments carrying therapeutic genes of interest expressed from native promoter sequences. This is of importance in situations where the therapy requires precise cell specific expression or in instances where expression is best achieved from genomic clones versus cDNA.

10 **3. Selectable markers for use, for example, in Gene Therapy (GT)**

 The following are selectable markers that can be incorporated into human *ACes* and used for selection.

15 **Dual Resistance to 4-Hydroperoxycyclophosphamide and Methotrexate by Retroviral Transfer of the Human Aldehyde Dehydrogenase Class 1 Gene and a Mutated Dihydrofolate Reductase Gene**

 The genetic transfer of drug resistance to hematopoietic cells is one approach to overcoming myelosuppression caused by high-dose chemotherapy. Because cyclophosphamide (CTX) and methotrexate (MTX) are commonly used non-cross-resistant drugs, generation of dual drug resistance in hematopoietic cells that allows dose intensification may increase anti-tumor effects and circumvent the emergence of drug-resistant tumors, a retroviral vector containing a human cytosolic ALDH-1-encoding DNA clone and a human doubly mutated DHFR-encoding clone (Phe22/Ser31; termed F/S in the description of constructs) to generate increased resistance to CTX and MTX were constructed (Takebe *et al.* (2001) *Mol Ther* 3(1):88-96). This construct may be useful for protecting patients from high-dose CTX- and MTX-induced myelosuppression. *ACes* can be similarly constructed.

Multiple mechanisms of N-phosphonacetyl-L-aspartate resistance in human cell lines: carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase gene amplification is frequent only when chromosome 2 is rearranged

5

Rodent cells resistant to N-phosphonacetyl-L-aspartate (PALA) invariably contain amplified carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase (CAD) genes, usually in widely spaced tandem arrays present as extensions of the same chromosome arm that carries a single copy of CAD in normal cells (Smith *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1816-21). In contrast, amplification of CAD is very infrequent in several human tumor cell lines. Cell lines with minimal chromosomal rearrangement and with unrearranged copies of chromosome 2 rarely develop intrachromosomal amplifications of CAD.

15 These cells frequently become resistant to PALA through a mechanism that increases the aspartate transcarbamylase activity with no increase in CAD copy number, or they obtain one extra copy of CAD by forming an isochromosome 2p or by retaining an extra copy of chromosome 2. In cells with multiple chromosomal aberrations and rearranged copies of chromosome 2, amplification of CAD as tandem arrays from rearranged chromosomes is the most frequent mechanism of PALA resistance. All of these different mechanisms of PALA resistance are blocked in normal human fibroblasts. Thus, ACes with multiple copies of the CAD gene would provide PALA resistance.

25

Retroviral coexpression of thymidylate synthase and dihydrofolate reductase confers fluoropyrimidine and antifolate resistance

Retroviral gene transfer of dominant selectable markers into hematopoietic cells can be used to select genetically modified cells in vivo or to attenuate the toxic effects of chemotherapeutic agents. Fantz *et al.* ((1998) *Biochem Biophys Res Comm* 243(1):6-12) have shown that

30

retroviral gene transfer of thymidylate synthase (TS) confers resistance to TS directed anticancer agents and that co-expression of TS and dihydrofolate reductase (DHFR) confers resistance to TS and DHFR cytotoxic agents. Retroviral vectors encoding *Escherichia coli* TS, human TS, and the Tyr-to-His at residue 33 variant of human TS (Y33HhTS) were constructed and fibroblasts transfected with these vectors conferred comparable resistance to the TS-directed agent fluorodeoxyuridine (FdUrd, approximately 4-fold). Retroviral vectors that encode dual expression of Y33HhTS and the human L22Y DHFR (L22YhDHFR) variants conferred resistance to FdUrd (3- to 5-fold) and trimetrexate (30- to 140-fold). A L22YhDHFR-Y33HhTS chimeric retroviral vector was also constructed and transduced cells were resistant to FdUrd (3-fold), AG337 (3-fold), trimetrexate (100-fold) and methotrexate (5-fold). These results show that recombinant retroviruses can be used to transfer the cDNA that encodes TS and DHFR and dual expression in transduced cells is sufficiently high to confer resistance to TS and DHFR directed anticancer agents. *ACes* can be similarly constructed.

Human CD34+ cells do not express glutathione S-transferases alpha

The expression of glutathione S-transferases alpha (GST alpha) in human hematopoietic CD34+ cells and bone marrow was studied using RT-PCR and immunoblotting (Czerwinski M, Kiem *et al.* (1997) *Gene Ther* 4(3):268-70). The GSTA1 protein conjugates glutathione to the stem cell selective alkylator busulfan. This reaction is the major pathway of elimination of the compound from the human body. Human hematopoietic CD34+ cells and bone marrow do not express GSTA1 message, which was present at a high level in liver, an organ relatively resistant to busulfan toxicity in comparison to bone marrow. Similarly, baboon CD34+ cells and dog bone marrow do not express GSTA1. Thus, human

GSTA1 is a chemoprotective selectable marker in human stem cell gene therapy and could be employed in *ACes* construction.

Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer

5 Pawliuk *et al.* ((1994) *Blood* 84(9):2868-2877) have investigated the use of a cell surface antigen as a dominant selectable marker to facilitate the detection and selection of retrovirally infected target cells. The small coding region of the human cell surface antigen CD24 (approximately 240 bp) was introduced into a myeloproliferative sarcoma virus (MPSV)-based retroviral vector, which was then used to infect day 4
10 5-fluorouracil (5-FU)-treated murine bone marrow cells. Within 48 hours of termination of the infection procedure CD24-expressing cells were selected by fluorescent-activated cell sorting (FACS) with an antibody directed against the CD24 antigen. Functional analysis of these cells
15 showed that they included not only *in vitro* clonogenic progenitors and day 12 colony-forming unit-spleen but also cells capable of competitive long-term hematopoietic repopulation. Double-antibody labeling studies performed on recipients of retrovirally transduced marrow cells showed that some granulocytes, macrophages, erythrocytes, and, to a lesser
20 extent, B and T lymphocytes still expressed the transduced CD24 gene at high levels 4 months later. No gross abnormalities in hematopoiesis were detected in mice repopulated with CD24-expressing cells. These results show that the use of the CD24 cell surface antigen as a retrovirally encoded marker permits rapid, efficient, and nontoxic selection *in vitro* of
25 infected primary cells, facilitates tracking and phenotyping of their progeny, and provides a tool to identify elements that regulate the expression of transduced genes in the most primitive hematopoietic cells. *ACes* could be similarly constructed.

DeltahGHR, a biosafe cell surface-labeling molecule for analysis and selection of genetically transduced human cells

A selectable marker for retroviral transduction and selection of human and murine cells is known (see, Garcia-Ortiz *et al.* (2000) *Hum Gene Ther* 11(2):333-46). The molecule expressed on the cell surface of the transduced population is a truncated version of human growth hormone receptor (deltahGHR), capable of ligand (hGH) binding, but devoid of the domains involved in signal triggering. The engineered molecule is stably expressed in the target cells as an inert protein unable to trigger proliferation or to rescue the cells from apoptosis after ligand binding. This new marker, has a wide application spectrum, since hGHR in the human adult is highly expressed only in liver cells, and lower levels have been reported in certain lymphocyte cell populations. The deltagHHR label has high biosafety potential, as it belongs to a well-characterized hormonal system that is nonessential in adults, and there is extensive clinical experience with hGH administration in humans. The differential binding properties of several monoclonal antibodies (MAbs) are used in a cell rescue method in which the antibody used to select deltagHHR-transduced cells is eluted by competition with hGH or, alternatively biotinylated hGH is used to capture tagged cells. In the latter system, the final purified population is recovered free of attached antibodies in hGH (a substance approved for human use)-containing medium. Such a system could be used to identify ACes containing cells.

4. Transgenic models for evaluation of genes and discovery of new traits in plants

Of interest is the use of plants and plant cells containing artificial chromosomes for the evaluation of new genetic combinations and discovery of new traits. Artificial chromosomes, by virtue of the fact that they can contain significant amounts of DNA can also therefore encode

numerous genes and accordingly a multiplicity of traits. It is contemplated here that artificial chromosomes, when formed from one plant species, can be evaluated in a second plant species. The resultant phenotypic changes observed, for example, can indicate the nature of the genes contained within the DNA contained within the artificial chromosome, and hence permit the identification of novel genetic activities. Artificial chromosomes containing euchromatic DNA or partially containing euchromatic DNA can serve as a valuable source of new traits when transferred to an alien plant cell environment. For example, it is contemplated that artificial chromosomes derived from dicot plant species can be introduced into monocot plant species by transferring a dicot artificial chromosome. The dicot artificial chromosome possessing a region of euchromatic DNA containing expressed genes.

The artificial chromosomes can be designed to allow the artificial chromosome to recombine with the naturally occurring plant DNA in such a fashion that a large region of naturally occurring plant DNA becomes incorporated into the artificial chromosome. This allows the artificial chromosome to contain new genetic activities and hence carry novel traits. For example, an artificial chromosome can be introduced into a wild relative of a crop plant under conditions whereby a portion of the DNA present in the chromosomes of the wild relative is transferred to the artificial chromosome. After isolation of the artificial chromosome, this naturally occurring region of DNA from the wild relative, now located on the artificial chromosome can be introduced into the domesticated crop species and the genes encoded within the transferred DNA expressed and evaluated for utility. New traits and gene systems can be discovered in this fashion. The artificial chromosome can be modified to contain sequences that promote homologous recombination within plant cells, or

be modified to contain a genetic system that functions as a site-specific recombination system.

Artificial chromosomes modified to recombine with plant DNA offer many advantages for the discovery and evaluation of traits in different
5 plant species. When the artificial chromosome containing DNA from one plant species is introduced into a new plant species, new traits and genes can be introduced. This use of an artificial chromosome allows for the ability to overcome the sexual barrier that prevents transfer of genes from one plant species to another species. Using artificial chromosomes in this
10 fashion allows for many potentially valuable traits to be identified including traits that are typically found in wild species. Other valuable applications for artificial chromosomes include the ability to transfer large regions of DNA from one plant species to another, such as DNA encoding potentially valuable traits such as altered oil, carbohydrate or protein
15 composition, multiple genes encoding enzymes capable of producing valuable plant secondary metabolites, genetic systems encoding valuable agronomic traits such as disease and insect resistance, genes encoding functions that allow association with soil bacterium such as growth promoting bacteria or nitrogen fixing bacteria, or genes encoding traits
20 that confer freezing, drought or other stress tolerances. In this fashion, artificial chromosomes can be used to discover regions of plant DNA that encode valuable traits.

The artificial chromosome can also be designed to allow the transfer and subsequent incorporation of these valuable traits now located
25 on the artificial chromosome into the natural chromosomes of a plant species. In this fashion the artificial chromosomes can be used to transfer large regions of DNA encoding traits normally found in one plant species into another plant species. In this fashion, it is possible to derive a plant cell that no longer needs to carry an artificial chromosome to

posses the novel trait. Thus, the artificial chromosome would serve as the transfer mechanism to permit the formation of plants with greater degree of genetic diversity.

The design of an artificial chromosome to accomplish the afore-
5 mentioned purposes can include within the artificial chromosome the presence of specific DNA sequences capable of acting as sites for homologous recombination to take place. For example, the DNA sequence of *Arabidopsis* is now known. To construct an artificial chromosome capable of recombining with a specific region of *Arabidopsis*
10 DNA, a sequence of *Arabidopsis* DNA, normally located near a chromosomal location encoding genes of potential interest can be introduced into an artificial chromosome by methods provided herein. It may be desirable to include a second region of DNA within the artificial chromosome that provides a second flanking sequence to the region
15 encoding genes of potential interest, to promote a double recombination event which would ensure transfer of the entire chromosomal region, encoding genes of potential interest, to the artificial chromosome. The modified artificial chromosome, containing the DNA sequences capable of homologous recombination region, can then be introduced into
20 *Arabidopsis* cells and the homologous recombination event selected.

It is convenient to include a marker gene to allow for the selection of a homologous recombination event. The marker gene is preferably inactive unless activated by an appropriate homologous recombination event. For example, US 5,272,071, describes a method where an
25 inactive plant gene is activated by a recombination event such that desired homologous recombination events can be easily scored. Similarly, US 5,501,967 describes a method for the selection of homologous recombination events by activation of a silent selection gene first introduced into the plant DNA, the gene being activated by an appropriate

homologous recombination event. Both of these methods can be applied to enable a selective process to be included to select for recombination between an artificial chromosome and a plant chromosome. Once the homologous recombination event is detected, the artificial chromosome, 5 once selected, is isolated and introduced into a recipient cell, for example, tobacco, corn, wheat or rice, and the expression of the newly introduced DNA sequences evaluated.

Phenotypic changes in the recipient plant cells containing the artificial chromosome, or in regenerated plants containing the artificial 10 chromosome, allows for the evaluation of the nature of the traits encoded by the *Arabidopsis* DNA, under conditions naturally found in plant cells, including the naturally occurring arrangement of DNA sequences responsible for the developmental control of the traits in the normal chromosomal environment.

15 Traits such as durable fungal or bacterial disease resistance, new oil and carbohydrate compositions, valuable secondary metabolites such as phytosterols, flavonoids, efficient nitrogen fixation or mineral utilization, resistance to extremes of drought, heat or cold are all found within different populations of plant species and are often governed by 20 multiple genes. The use of single gene transformation technologies does not permit the evaluation of the multiplicity of genes controlling many valuable traits. Thus, incorporation of these genes into artificial chromosomes allows the rapid evaluation of the utility of these genetic combinations in heterologous plant species.

25 The large scale order and structure of the artificial chromosome provides a number of unique advantages in screening for new utilities or novel phenotypes within heterologous plant species. The size of new DNA that can be carried by an artificial chromosome can be millions of base pairs of DNA, representing potentially numerous genes that may

have novel utility in a heterologous plant cell. The artificial chromosome is a "natural" environment for gene expression, the problems of variable gene expression and silencing seen for genes transferred by random insertion into a genome should not be observed. Similarly, there is no
5 need to engineer the genes for expression, and the genes inserted would not need to be recombinant genes. Thus, one expects the expression from the transferred genes to be temporal and spatial, as observed in the species from where the genes were initially isolated. A valuable feature for these utilities is the ability to isolate the artificial chromosomes and to
10 further isolate, manipulate and introduce into other cells artificial chromosomes carrying unique genetic compositions.

Thus, the use of artificial chromosomes and homologous recombination in plant cells can be used to isolate and identify many valuable crop traits.

15 In addition to the use of artificial chromosomes for the isolation and testing of large regions of naturally occurring DNA, methods for the use of artificial chromosomes and cloned DNA are also contemplated. Similar to that described above, artificial chromosomes can be used to carry large regions of cloned DNA, including that derived from other plant species.

20 The ability to incorporate novel DNA elements into an artificial chromosome as it is being formed allows for the development of artificial chromosomes specifically engineered as a platform for testing of new genetic combinations, or "genomic" discoveries for model species such as *Arabidopsis*. It is known that specific "recombinase" systems can be
25 used in plant cells to excise or re-arrange genes. These same systems can be used to derive new gene combinations contained on an artificial chromosome.

The artificial chromosomes can be engineered as platforms to accept large regions of cloned DNA, such as that contained in Bacterial

Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs).

It is further contemplated, that as a result of the typical structure of artificial chromosomes containing tandemly repeated DNA blocks, that sequences other than cloned DNA sequence can be introduced by

5 recombination processes. In particular recombination within a predefined region of the tandemly repeated DNA within the artificial chromosome provides a mechanism to "stack" numerous regions of cloned DNA, including large regions of DNA contained within BACs or YACs clones. Thus, multiple combinations of genes can be introduced onto artificial
10 chromosomes and these combinations tested for functionality. In particular, it is contemplated that multiple YACs or BACs can be stacked onto an artificial chromosomes, the BACs or YACs containing multiple genes of complex pathways or multiple genetic pathways. The BACs or YACs are typically selected based on genetic information available within
15 the public domain, for example from the Arabidopsis Information Management System (<http://aims.cps.msu.edu/aims/index.html>) or the information related to the plant DNA sequences available from the Institute for Genomic Research (<http://www.tigr.org>) and other sites known to those skilled in the art. Alternatively, clones can be chosen at
20 random and evaluated for functionality. It is contemplated that combinations providing a desired phenotype can be identified by isolation of the artificial chromosome containing the combination and analyzing the nature of the inserted cloned DNA.

In this regard, it is contemplated that the use of site-specific
25 recombination sequences can have considerable utility in developing artificial chromosomes containing DNA sequences recognized by recombinase enzymes and capable of accepting DNA sequences containing same. The use of site-specific recombination as a means to target an introduced DNA to a specific locus has been demonstrated in

the art and such methods can be employed. The recombinase systems can also be used to transfer the cloned DNA regions contained within the artificial chromosome to the naturally occurring plant or mammalian chromosomes.

5 As noted herein, many site-specific recombinases are known and can be identified (Kilby *et al.* (1993) *Trends in Genetics* 9:413-418). The three recombinase systems that have been extensively employed include: an activity identified as R encoded by the pSR1 plasmid of *Zygosaccharomyces rouxii*, FLP encoded for the 2 μ m circular plasmid from
10 *Saccharomyces cerevisiae* and *Cre-lox* from the phage P1.

 The integration function of site-specific recombinases is contemplated as a means to assist in the derivation of genetic combinations on artificial chromosomes. In order to accomplish this, it is contemplated that a first step of introducing site-specific recombinase
15 sites into the genome of a plant cell in an essentially random manner is conducted, such that the plant cell has one or more site-specific recombinase recognition sequences on one or more of the plant chromosomes. An artificial chromosome is then introduced into the plant cell, the artificial chromosome engineered to contain a recombinase
20 recognition site (e.g., integration site) capable of being recognized by a site-specific recombinase. Optionally, a gene encoding a recombinase enzyme is also included, preferably under the control of an inducible promoter. Expression of the site-specific recombinase enzyme in the plant cell, either by induction of a inducible recombinase gene, or
25 transient expression of a recombinase sequence, causes a site-specific recombination event to take place, leading to the insertion of a region of the plant chromosomal DNA (containing the recombinase recognition site) into the recombinase recognition site of the artificial chromosome, and forming an artificial chromosome containing plant chromosomal DNA.

The artificial chromosome can be isolated and introduced into a heterologous host, preferably a plant host, and expression of the newly introduced plant chromosomal DNA can be monitored and evaluated for desirable phenotypic changes. Accordingly, carrying out this

5 recombination with a population of plant cells wherein the chromosomally located recombinase recognition site is randomly scattered throughout the chromosomes of the plant, can lead to the formation of a population of artificial chromosomes, each with a different region of plant chromosomal DNA, and each potentially representing a novel genetic combination.

10 This method requires the precise site-specific insertion of chromosomal DNA into the artificial chromosome. This precision has been demonstrated in the art. For example, Fukushige and Sauer ((1992) Proc. Natl. Acad. Sci. USA, 89:7905-7909) demonstrated that the *Cre-lox* homologous recombination system could be successfully employed to
15 introduce DNA into a predefined locus in a chromosome of mammalian cells. In this demonstration a promoter-less antibiotic resistance gene modified to include a *lox* sequence at the 5' end of the coding region was introduced into CHO cells. Cells were re-transformed by electroporation with a plasmid that contained a promoter with a *lox* sequence and a
20 transiently expressed *Cre* recombinase gene. Under the conditions employed, the expression of the *Cre* enzyme catalyzed the homologous recombination between the *lox* site in the chromosomally located promoter-less antibiotic resistance gene, and the *lox* site in the introduced promoter sequence, leading to the formation of a functional antibiotic
25 resistance gene. The authors demonstrated efficient and correct targeting of the introduced sequence, 54 of 56 lines analyzed corresponded to the predicted single copy insertion of the DNA due to *Cre* catalyzed site-specific homologous recombination between the *lox* sequences.

Accordingly a *lox* sequence may be first added to a genome of a plant species capable of being transformed and regenerated to a whole plant to serve as a recombinase target DNA sequence for recombination with an artificial chromosome. The *lox* sequence may be optimally
5 modified to further contain a selectable marker which is inactive but can be activated by insertion of the *lox* recombinase recognition sequence into the artificial chromosome.

A promoterless marker gene or selectable marker gene linked to the recombinase recognition sequence, which is first inserted into the
10 chromosomes of a plant cell can be used to engineer a platform chromosome. A promoter is linked to a recombinase recognition site, in an orientation that allows the promoter to control the expression of the marker or selectable marker gene upon recombination within the artificial chromosome. Upon a site-specific recombination event between a
15 recombinase recognition site in a plant chromosome and the recombinase recognition site within the introduced artificial chromosome, a cell is derived with a recombined artificial chromosome, the artificial chromosome containing an active marker or selectable marker activity that permits the identification and or selection of the cell.

20 The artificial chromosomes can be transferred to other plant or animal species and the functionality of the new combinations tested. The ability to conduct such an inter-chromosomal transfer of sequences has been demonstrated in the art. For example, the use of the *Cre-lox* recombinase system to cause a chromosome recombination event
25 between two chromatids of different chromosomes has been shown.

Any number of recombination systems may be employed as described herein, such as, but not limited to, bacterially derived systems such as the att/int system of phage lambda, and the Gin/gix system.

More than one recombination system may be employed, including, for example, one recombinase system for the introduction of DNA into an artificial chromosome, and a second recombinase system for the subsequent transfer of the newly introduced DNA contained within an
5 artificial chromosome into the naturally occurring chromosome of a second plant species. The choice of the specific recombination system used will be dependent on the nature of the modification contemplated.

By having the ability to isolate an artificial chromosome, in particular, artificial chromosomes containing plant chromosomal DNA
10 introduced via site-specific recombination, and re-introduce the chromosome into other mammalian or plant cells, particularly plant cells, these new combinations can be evaluated in different crop species without the need to first isolate and modify the genes, or carry out multiple transformations or gene transfers to achieve the same
15 combination isolation and testing combinations of the genes in plants. The use of a site-specific recombinase also allows the convenient recovery of the plant chromosomal region into other recombinant DNA vectors and systems, such as mammalian or insect systems, for manipulation and study.

20 Also contemplated herein are *ACes*, cell lines and methods for use in screening a new chromosomal combinations, deletions, truncations with eucaryotic genome that take advantage of the site-specific recombination systems incorporated onto platform *ACes* provided herein. For example, provided herein is a cell line useful for making a library of
25 *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes. Also provided herein is a method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into a cell line comprising a multiplicity of heterologous recombination

- sites randomly integrated throughout the endogenous chromosomes, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites within the cell's chromosomal DNA; and isolating said
- 5 multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within. Also provided herein is a library of cells useful for genomic screening, said library comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein. The
- 10 library of cells can be from a different species and/or cell type than the chromosomal nucleic acid within the *ACes*. Also provided is a method of making one or more cell lines, comprising
- a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,
 - 15 b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;
 - c) isolating said multiplicity of *ACes*, thereby producing a library of
 - 20 *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;
 - d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;
 - e) selecting different cells having mutually exclusive *ACes* therein
 - 25 and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

These *ACes*, cell lines and methods utilize the site-specific recombination sites on platform *ACes* analogous YAC manipulation related to: the methods of generating terminal deletions in normal and

artificial chromosomes (e.g., *ACes*; as described in Vollrath et al., 1988, *PNAS, USA*, 85:6027-66031; and Pavan et al., *PNAS, USA*, 87:1300-1304); the methods of generating interstitial deletions in normal and artificial chromosomes (as described in Campbell et al., 1991, *PNAS, USA*, 88:5744-5748); and the methods of detecting homologous recombination between two *ACes* (as described in Cellini et al., 1991, *Nuc. Acid Res.*, 19(5):997-1000).

5. Use of platform *ACes* in Pharmacogenomic/toxicology applications (development of "Reporter *ACes*")

- 10 In addition to the placement of genes onto *ACes* chromosomes for therapeutic protein production or gene therapy, the platform can be engineered via the IntR lambda integrase to carry reporter-linked constructs (reporter genes) that monitor changes in cellular physiology as measured by the particular reporter gene (or a series of different reporter genes) readout.
- 15 The reporter linked constructs are designed to include a gene that can be detected (by for example fluorescence, drug resistance, immunohistochemistry, or transcript production, and the like) with well-known regulatory sequences that would control the expression of the detectable gene. Exemplary regulatory promoter sequences are well-known
- 20 in the art.

A) Reporter *ACes* for drug pathway screening

- The *ACes* can be engineered to carry reporter-linked constructs that indicate a signal is being transduced through one or a number of pathways. For example, transcriptionally regulated promoters from genes at the end (or
- 25 any other chosen point) of particular signal transduction pathways could be engineered on the *ACes* to express the appropriate readout (either by fluorescent protein production or drug resistance) when the pathway is activated (or down-regulated as well). In one embodiment, a number of reporters from different pathways can be placed on an

ACes chromosome. Cells (and/or whole animals) containing such a Reporter ACes could be exposed to a variety of drugs or compounds and monitored for the effects of the drugs or compounds upon the selected pathway(s) by the reporter gene(s). Thus, drugs or compounds can be
5 classified or identified by particular pathways they excite or down-regulate. Similarly, transcriptional profiles obtained from genomic array experiments can be biologically validated using the reporter ACes provided herein.

B) Reporter ACes for toxic compound testing

10 Environmental or man-made genotoxicants can be tested in cell lines carrying a number of reporter-genes platform ACes linked to promoters that are transcriptionally regulated in response to DNA damage, induced apoptosis or necrosis, and cell-cycle perturbations. Furthermore, new drugs and/or compounds could be tested in a similar manner with the
15 genotoxicant ACes reporter for their cellular/genetic toxicity by such a screen. Likewise, toxic compound testing could be carried out in whole transgenic animals carrying the ACes chromosome that measures genotoxicant exposure ("canary in a coal mine"). Thus, the same or similar type ACes could be used for toxicity testing in either a cell-based
20 or whole animal setting. An example would include ACes that carry reporter-linked genes controlled by various cytochrome P450 profiled promoters and the like.

C) Reporter ACes for individualized pharmacogenomics/drug profiling

25 A common disease may arise via various mechanisms. In many instances there are multiple treatments available for a given disease. However, the success of a given treatment may depend upon the mechanism by which the disease originated and/or by the genetic background of the patient. In order to establish the most effective

treatment for a given patient one could utilize the *ACes* reporters provided herein. *ACes* reporters can be used in patient cell samples to determine an individualized drug regimen for the patient. In addition, potential polymorphisms affecting the transcriptional regulation of an individual's particular gene can be assessed by this approach.

D) Reporter *ACes* for classification of similar patient tumors

As with other diseases as described in 5.C) above, cancer cells arise via different mechanisms. Furthermore, as a cancerous cell propagates it may undergo genomic alterations. An *ACes* reporter transferred to cells of different patients having the same disease, i.e. similar cancers, could be used to categorize the particular cancer of each patient, thereby facilitating the identification of the most effective therapeutic regimen. Examples would include the validation of array profiling of certain classes of breast cancers. Subsequently, appropriate drug profiling could be carried out as described above.

E) Reporter *ACes* as a "differentiation" sensor

Using the *ACes* reporter as a "differentiation" sensor in stem cells or other progenitor cells in order to enrich by selection (either FACS based screening, drug selection and/or use of suicide gene) for a particular class of differentiated or undifferentiated cells. For example, in one embodiment, this assay could also be used for compound screening for small molecule modifiers of cell differentiation.

F) Whole animal studies with Reporter *ACes*

Finally, with whole-body fluorescence imaging technology (Yang et al. (2000) PNAS 97:12278) any of the above Reporter *ACes* methods could be used in conjunction with whole-body imaging to monitor reporter genes within whole animals without sacrificing the animal. This would allow temporal and spatial analysis of expression patterns under a given set of conditions. The conditions tested may include for example, normal

differentiation of a stem cell, response to drug or compound treatment whether targeted to the diseased tissue or presented systemically, response to genotoxicants, and the like.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLE 1

pFK161

Cosmid pFK161 (SEQ ID NO: 118) was obtained from Dr. Gyula Hadlaczky and contains a 9 kb *NotI* insert derived from a murine rDNA repeat (see clone 161 described in PCT Application Publication No. WO97/40183 by Hadlaczky *et al.* for a description of this cosmid). This cosmid, referred to as clone 161 contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 26. It was produced by inserting fragments of the megachromosome (see, U.S. Patent No. 6,077,697 and International PCT application No. WO 97/40183). For example, H1D3, which was deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, is a mouse-hamster hybrid cell line carrying this megachromosome into plasmid pWE15 (Stratagene, La Jolla, California; SEQ ID No. 31) as follows. Half of a 100 μ l low melting point agarose block (mega-plug) containing isolated SATACs was digested with *NotI* overnight at 37°C. Plasmid pWE15 was similarly digested with *NotI* overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 DNA ligase. Ligation was conducted at 16°C overnight. Bacterial DH5 α cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and analyzed by Southern blot hybridization for the presence of DNA that

hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected.

Liquid cultures of all 189 transformants were used to generate
5 cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones
10 (113, 157 and 161) contained the same insert.

For sequence analysis the insert of cosmid clone no. 161 was subcloned as follows. To obtain the end fragments of the insert of clone no. 161, the clone was digested with *NotI* and *Bam*HI and ligated with *NotI*/*Bam*HI-digested pBluescript KS (Stratagene, La Jolla, California).

15 Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with *Bam*HI-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

20 The insert corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 18. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 19-25. Specifically, the individual
25 subclones corresponded to the following positions in GENBANK accession no. X82564 (SEQ ID NO:18) and in SEQ ID NOs. 19-25:

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<i>NotI, BamHI</i>	19
161m5	7756	8494	<i>BamHI</i>	20
161m7	8495	10231	<i>BamHI</i>	21 (shows only sequence corresponding to nt. 8495-8950), 22 (shows only sequence corresponding to nt. 9851- 10231)
161m12	10232	15000	<i>BamHI</i>	23 (shows only sequence corresponding to nt. 10232-10600), 24 (shows only sequence corresponding to nt. 14267-15000)
161k2	15001	15676	<i>NotI, BamHI</i>	25

5 The sequence set forth in SEQ ID NOs. 19-25 diverges in some
10 positions from the sequence presented in positions 7551-15670 of
GENBANK accession no. X82564. Such divergence may be attributable
to random mutations between repeat units of rDNA.

For use herein, the rDNA insert from the clone was prepared by
digesting the cosmid with *NotI* and *BglII* and was purified as described
15 above. Growth and maintenance of bacterial stocks and purification of
plasmids were performed using standard well known methods (see, *e.g.*,
Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd
Edition, Cold Spring Harbor Laboratory Press), and plasmids were purified
from bacterial cultures using Midi - and Maxi-preps Kits (Qiagen,
20 Mississauga, Ontario).

pDsRed1N1

This vector is available from Clontech (see SEQ ID No. 29) and
encodes the red fluorescent protein (DsRed; Genbank accession no.
AF272711; SEQ ID Nos. 39 and 40). DsRed, which has a vivid red
25 fluorescence, was isolated from the IndoPacific sea anemone relative
Discosoma species. The plasmid pDsRed1N1 (Clontech; SEQ ID No. 29)
constitutively expresses a human codon-optimized variant of the

fluorescent protein under control of the CMV promoter. Unmodified, this vector expresses high levels of DsRed1 and includes sites for creating N-terminal fusions by cloning proteins of interest into the multiple cloning site (MCS). It is Kan and Neo resistant for selection in bacterial or
5 eukaryotic cells.

Plasmid pMG

Plasmid pMG (InvivoGen, San Diego, California; see SEQ. ID. NO. 27 for the nucleotide sequence of pMG) contains the hygromycin phosphotransferase gene under the control of the immediate-early human
10 cytomegalovirus (hCMV) enhancer/promoter with intron A. Vector pMG also contains two transcriptional units allowing for the coexpression of two heterologous genes from a single vector sequence.

The first transcriptional unit of pMG contains a multiple cloning site for insertion of a gene of interest, the hygromycin phosphotransferase
15 gene (*hph*) and the immediate-early human cytomegalovirus (hCMV) enhancer/promoter with intron A (see, *e.g.*, Chapman *et al.* (1991) *Nuc. Acids Res.* 19:3979-3986) located upstream of *hph* and the multiple cloning site, which drives the expression of *hph* and any gene of interest inserted into the multiple cloning site as a polycistronic mRNA. The first
20 transcriptional unit also contains a modified EMCV internal ribosomal entry site (IRES) upstream of the *hph* gene but downstream of the hCMV promoter and MCS for ribosomal entry in translation of the *hph* gene (see SEQ ID NO. 27, nucleotides 2736-3308). The IRES is modified by insertion of the constitutive *E. coli* promoter (EM7) within an intron (IM7)
25 into the end of the IRES. In mammalian cells, the *E. coli* promoter is treated as an intron and is spliced out of the transcript. A polyadenylation signal from the bovine growth hormone (bGh) gene (see, *e.g.*, Goodwin and Rottman (1992) *J. Biol. Chem.* 267:16330-16334) and a pause site derived from the 3' flanking region of the human $\alpha 2$

globin gene (see, *e.g.*, Enriquez-Harris *et al.* (1991) *EMBO J.* 10:1833-1842) are located at the end of the first transcription unit. Efficient polyadenylation is facilitated by inserting the flanking sequence of the bGh gene 3' to the standard AAUAAA hexanucleotide sequence.

5 The second transcriptional unit of pMG contains another multiple cloning site for insertion of a gene of interest and an EF-1 α /HTLV hybrid promoter located upstream of this multiple cloning site, which drives the expression of any gene of interest inserted into the multiple cloning site. The hybrid promoter is a modified human elongation factor-1 alpha (EF-1
10 alpha) gene promoter (see, *e.g.*, Kim *et al.* (1990) *Gene* 91:217-223) that includes the R segment and part of the U5 sequence (R-U5') of the human T-cell leukemia virus (HTLV) type I long terminal repeat (see, *e.g.*, Takebe *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). The Simian Virus 40 (SV40) late polyadenylation signal (see Carswell and Alwine (1989) *Mol.*
15 *Cell. Biol.* 9:4248-4258) is located downstream of the multiple cloning site. Vector pMG contains a synthetic polyadenylation site for the first and second transcriptional units at the end of the transcriptional unit based on the rabbit β -globin gene and containing the AATAAA hexanucleotide sequence and a GT/T-rich sequence with 22-23
20 nucleotides between them (see, *e.g.*, Levitt *et al.* (1989) *Genes Dev.* 3:1019-1025). A pause site derived from the C2 complement gene (see, Moreira *et al.* (1995) *EMBO J.* 14:3809-3819) is also located at the 3' end of the second transcriptional unit.

 Vector pMG also contains an ori sequence (ori pMB1) located
25 between the SV40 polyadenylation signal and the synthetic polyadenylation site.

EXAMPLE 2

A. Construction of targeting vector and transfection into LMtk- cells for the generation of platform chromosomes

A targeting vector derived from the vector pWE15 (GeneBank Accession # X65279) was modified by replacing the *Sa*I (Klenow filled)/*Sma*I neomycin resistance containing fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance containing fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc. Palo Alto, CA; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (SEQ ID NO: 118) containing a portion of the mouse rDNA region was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K (Figure 2). The vector pWEPuro9K was digested with *Spe*I to linearize and transfected into LMtk- mouse cells. Puromycin resistant colonies were isolated and subsequently tested for artificial chromosome formation via fluorescent *in situ* hybridization (FISH) (using mouse major and minor DNA repeat sequences, the puromycin gene and telomeres sequences as probes), and fluorescent activated cell sorting (FACS). From this sort, a subclone was isolated containing an artificial chromosome, designated 5B11.12, which carries 4-8 copies of the puromycin resistance gene contained on the pWEPuro9K vector. FISH analysis of the 5B11.12 subclone demonstrated the presence of telomeres and mouse minor on the ACes. DOT PCR has been done on the 5B11.12 ACes revealing the absence of uncharacterized euchromatic regions on the ACes. A recombination site, such as an *att* or *loxP* engineering site or a plurality thereof, was introduced onto this ACes thereby providing a platform for site-specific introduction of heterologous nucleic acid.

B. Targeting a single sequence specific recombination site onto platform chromosomes

After the generation of the 5B11.12 platform, a single sequence-specific recombination site is placed onto the platform chromosome via homologous recombination. For this, DNA sequences containing the site-

specific recombination sequence can be flanked with DNA sequences of homology to the platform chromosome. For example, using the platform chromosome made from the pWEPuro9K vector, mouse rDNA sequences or mouse major satellite DNA can be used as homologous sequences to target onto the platform chromosome. A vector is designed to have these homologous sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the platform chromosome, the DNA is transfected into cells harboring the platform chromosome (Figure 3). Examples of site-specific cassettes that are targeted to the platform chromosome using either mouse rDNA or mouse major repeat DNA include the SV40-attP-hygro cassette and a red fluorescent protein (RFP) gene flanked by loxP sites (Cre/lox, see, *e.g.*, U.S. Patent No. 4,959,317 and description herein). After transfection and integration of the site-specific cassette, homologous recombination events onto the platform chromosome are subcloned and identified by FACS (e.g. screen and single cell subclone via expression of resistance or fluorescent marker) and PCR analysis.

For example, a vector can be constructed containing regions of the mouse rDNA locus flanking a gene cassette containing the SV40 early reporter-bacteriophage lambda attP site-hygromycin selectable marker (see Figure 4 and described below). The use of the bacteriophage lambda attP site for lambda integrase-mediated site-specific recombination is described below. Homologous recombination event of the SV40-attP-hygro cassette onto the platform chromosome was identified using PCR primers that detect the homologous recombination and further confirmed by FISH analysis. After identifying subcloned colonies containing the platform chromosome with a single site-specific recombination site, cells carrying the platform chromosome with a single site-specific

recombination site can now be engineered with site-specific recombinases (e.g. lambda INT, Cre) for integrating a target gene expression vector.

C. Targeting a red fluorescent protein (RFP) gene flanked by loxP sites onto 5B11.12 platform

As another example, while loxP recombination sites could have been introduced onto the *ACes* during *de novo* biosynthesis, it was thought that this might result in multiple segments of the *ACes* containing a high number of loxP sites, potentially leading to instability upon *Cre*-mediated recombination. A gene targeting approach was therefore devised to introduce a more limited number of loxP recombination sites into a locus of the 5B11-12 *ACes* containing introduced and possibly co-amplified endogenous rDNA sequences. Although there are more than 200 copies of rDNA genes in the haploid mouse genome distributed amongst 5-11 chromosomes (depending on strain), rDNA sequences were chosen as the target on the *ACes* since they represent a less frequent target than that of the satellite repeat sequences. Moreover, having observed much stronger pWEPuro9K hybridization to the 5B11-12 *ACes* than to other LMTK⁻ chromosomes and in light of the observation that the transcribed spacer sequences within the rDNA may be less conserved than the rRNA coding regions, it was contemplated that a targeting vector based on the rDNA gene segment in pWEPuro9K would have a higher probability of targeting to the *ACes* rather than to other LMTK⁻ chromosomes. Accordingly, a targeting vector, pBSFKLoxDsRedLox, was designed and constructed based on the rDNA sequences contained in pWEPuro9K.

The plasmid pBSFKLoxDsRedLox was generated in 4 steps. First, the *NotI* rDNA insert of pWEPuro9K (Figure 2) was inserted into pBS SK- (Stratagene) giving rise to pBSFK. Second, a loxP polylinker cassette was

generated by PCR amplification of pNEB193 (SEQ ID NO:32; New England Biolabs) using primers complementary to the M13 forward and reverse priming sites at their 3' end and a 34 bp 5' extension comprising a LoxP site. This cassette was reinserted into pNEB193 generating

5 p193LoxMCSLox. Third, the DsRed gene from pDsRed1-N1 (SEQ ID NO:29; Clontech) was then cloned into the polylinker between the loxP sites generating p193LoxDsRedLox. Fourth, a fragment consisting of the DsRed gene flanked by loxP sites was cloned into a unique *NdeI* within the rDNA insert of pBSFK generating pBSFKLoxDsRedLox.

10 A gel purified 11 Kb *PmlI* /*EcoRV* fragment of pBSFKLoxDsRedLox was used for transfection. To detect targeted integration, PCR primers were designed from rDNA sequences within the 5' *NotI*-*PmlI* fragment of pWEPuro9K that is not present on the targeting fragment (5' primer) and sequence within the LoxDsRedLox cassette (3' primer). If the targeting
15 DNA integrated correctly within the rDNA sequences, PCR amplification using these primers would give rise to a 2.3 Kb band. PCR reactions containing 1-4 μ l of genomic DNA were carried out according to the MasterTaq protocol (Eppendorf), using murine rDNA 5' primer (5'-CGGACAATGCGGTTGTGCGT-3'; SEQ ID NO:72) and DsRed 3' primer
20 (5'GGCCCCGTAATGCAGAAGAA-3'; SEQ ID NO:73) and PCR products were analyzed by agarose gel electrophoresis.

1.5X10⁶ 5B11-12 LMTK⁻ cells were transfected with 2 μ g of the pBSFKLoxDsRedLox targeting DNA described above using Lipofectamine Plus (Invitrogen). For flow sorting, harvested cells were suspended in
25 medium and applied to the Becton Dickinson Vantage SE cell sorter, equipped with 488 nm lasers for excitation and 585/42 bandpass filter for optimum detection of RFP fluorescence. Cells were sorted using dPBS as sheath buffer. Negative control parental 5B11-12 cells and a positive control LMTK⁻ cell line stably transfected with DsRed were used to

establish the selection gates. The RFP positive gated populations were recovered, diluted in medium supplemented with 1X penicillin-streptomycin (Invitrogen), then plated and cultured as previously described. After 4 rounds of enrichment, the percentage of RFP positive cells reached levels of 50% or higher. DNA from populations was analyzed by PCR for evidence of targeted integration. Ultimately, single cell subclones were established from positive pools and were analyzed by PCR and PCR-positive clones confirmed by FISH as described below. DNA was purified from pools or single cell clones using previously described methods set forth in Lahm et al., Transgenic Res., 1998; 7:131-134, or in some cases using a Wizard Genomic DNA purification kit (Promega). For FISH analysis, a biotinylated DsRed gene probe was generated by PCR using DsRed specific primers and biotin-labeled dUTP (5' RFP primer: 5'-GGTTTAAAGTGCGCTCCTCCAAGAACGTCATC-3', SEQ ID NO:74; and 3' RFP primer: 5'AGATCTAGAGCCGCGCTACAGGAACAGGTGGTGGCGGCC-3'; SEQ ID NO:75). To maximize the signal intensity of the DsRed probe, Tyramide amplification was carried out according to the manufacturers protocols (NEN).

20 The process of testing the feasibility of a more general targeting strategy that would not rely on enrichment *via* drug selection of stably transfected clones can be summarized as follows. A red fluorescent protein gene (RFP; encoded by the DsRed gene) was inserted between the loxP sites of the targeting vector to form pBSFKLoxDsRedLox. After

25 transfection with pBSFKLoxDsRedLox, sequential rounds of high speed flow sorting and expansion of sorted cells in culture could then be used to enrich for stable transformants expressing RFP. In the event of targeted integration, PCR screening with primers that amplify from a spacer region within the segment of the 45s pre-rRNA gene in pWEPuro9K to a specific

anchor sequence within the DsRed gene in the targeting cassette would give rise to a diagnostic 2.3 Kb band. However, as rDNA clusters are found on several chromosomes, confirmation of targeting to an *ACes* would require fluorescence in situ hybridization (FISH) analysis. Finally, the flanking of the DsRed gene by loxP sites would allow for its removal and subsequent replacement with other genes of interest.

After transfection of the targeting sequence into 5B11-12 cells, enrichment for targeted clones was carried out using a combination of flow cytometry to detect red-fluorescing cells and PCR screening.

10 Ultimately 17 single cell subclones were identified as potential targeted clones by PCR and of these 16 were found by FISH to contain the DsRed integration event into the *ACes*. These subclones are referred to herein as D11-C4, D11-C12, D11-H3, C9-C9, C9-B9, C9-F4, C9-H8, C9-F2, C9-G8, C9-B6, C9-G3, C9-E12, C9-A11, C11-E3, C11-A9 and C11-H4. PCR

15 analysis of genomic DNA isolated from the D11-C4 subclone gave rise to a 2.3 Kb band, indicative of a targeted integration into an rDNA locus. Further analysis of the subclone by FISH analysis with a DsRed gene probe demonstrated integration of the LoxDsRedLox targeting cassette on the *ACes* co-localizing with one of the regions of rDNA staining seen on

20 the 5B11-12 *ACes*, consistent with a targeted integration into an rDNA locus of the *ACes*, while integrations on other chromosomes were not observed. Since transfected cells were maintained as heterogeneous populations through several cycles of sorting and replating it was not possible to estimate the frequency of targeted events. In most

25 mammalian cell lines the frequency of gene targeting via homologous recombination is roughly 10^{-5} - 10^{-7} treated cells. Despite the low frequency of these events in mammalian cells, it is clear that an RFP expression based screening paradigm, coupled with PCR analysis, can effectively detect and enrich for such infrequent events in a large

population. In instances where drug selection is not possible or not desirable, such a system may provide a useful alternative. It was also verified that the modified *ACes* in subclone D11-C4 could be purified by flow cytometry. The results indicate that the flow karyogram of the D11-C4 subclone was unaltered from that of the 5B11-12 cell line. Thus, the D11-C4 *ACes* can be purified in high yield from native chromosomes of the host cell line.

D. Reduction of LoxP on *ACes* to a single site.

10 The strong hybridization signal detected by FISH on the *ACes* using the DsRed gene probe suggests that several copies of the targeting cassette may be present on the *ACes* in the D11-C4 line. This also suggests that multiple rDNA genes have been correctly targeted.

15 Accordingly, in certain embodiments where necessary, the number of loxP sites on the *ACes* can be reduced to a single site by *in situ* treatment with *Cre* recombinase, provided that the sites are co-linear. Such a process is described for multiple loxP-flanked integrations on a native mouse chromosome (Garrick et al., Nature Genet., 1998, Jan;18(1):56-59). Reduction to a single loxP site on the D11-C4 *ACes* would result in the loss of the DsRed gene, forming the basis of a useful screen for this event.

25 For this purpose, a Cre expression plasmid pCX-Cre/GFP III has been generated by first deleting the EcoRI fragment of pCX-eGFP (SEQ ID NO:71) containing the eGFP coding sequence and replacing it with that of a PCR amplified Cre recombinase coding sequence (SEQ ID NO:58), generating pCX-Cre. Next, the Asel/Sspl fragment of pD2eGFP-N1 (containing the CMV promoter driving the D2EGFP gene with SV40 polyA signal; Clontech; SEQ ID NO:87) was inserted into the filled HindIII site of pCX-Cre, generating pCX-Cre\GFP III. Control plasmid pCX-CreRev\GFP

III was generated in similar fashion except that the Cre recombinase coding sequence was inserted in the antisense orientation. LMTK⁻ cell line D11-C4 (containing first generation platform ACes with multiple loxP-DsRED sites) and 5B11-12 cell line (containing ACes with no loxP-DsRED sites) are maintained in culture as described above. D11C4 cells are transfected with 2 μ g of plasmid pCX-Cre\GFP III or 2 μ g pCX-CreRev\GFP III using Lipofectamine (Invitrogen) as previously described.

Forty-eight to seventy-two hours after transfection, transfected D11-C4 cells are harvested and GFP positive cells are sorted by cell cytometry using a FACSta Vantage cell sorter (Beckton-Dickinson) as follows: All D11-C4 cells transfected with pCX-Cre\GFP III or control plasmid pCX-CreRev\GFP III that exhibit GFP fluorescent higher than the gate level established by untransfected cells are collected and placed in culture a further 7-14 days. After 7-14 days the initial D11-C4 cells are harvested and analyzed by cell cytometry as follows: Untransfected D11-C4 cells are used to establish the gate that defines the RFP positive population, while 5B11-12 cells are used to set the RFP negative gate. The GFP positive population of D11-C4 transfected with pCX-Cre\GFP III should show decreased red fluorescence compared to pCX-CreRev\GFP III transfected or untransfected control D11-C4 cells. The cells exhibiting greatly decreased or no RFP expression are collected and single cell clones subsequently established. These clones will be expanded and analyzed by fluorescence *in-situ* hybridization and Southern blotting to confirm the removal of loxP-DsRed gene copies.

25

EXAMPLE 3

Construction of targeting vector and transfection into LMTK⁻ cells for the generation of platform chromosomes containing multiple site-specific recombination sites

An example of a selectable marker system for the creation of a chromosome-based platform is shown in Figure 4. This system includes a vector containing the SV40 early promoter immediately followed by (1) a 282 base pair (bp) sequence containing the bacteriophage lambda attP site and (2) the puromycin resistance marker. Initially a *PvuII/StuI* fragment containing the SV40 early promoter from plasmid pPUR (Clontech Laboratories, Inc., Palo Alto, CA; Seq ID No. 30) was subcloned into the *EcoRI/CRI* site of pNEB193 (a PUC19 derivative obtained from New England Biolabs, Beverly, MA; SEQ ID No. 32) generating the plasmid pSV40193. The only differences between pUC19 and pNEB193 are in the polylinker region. A unique *Ascl* site (GGCGCGCC) is located between the *BamHI* site and the *SmaI* site, a unique *PacI* site (TTAATTAA) is located between the *BamHI* site and the *XbaI* site and a unique *PmeI* site (GTTTAAAC) is located between the *PstI* site and the *SaI* site.

The attP site was PCR amplified from lambda genome (GenBank Accession # NC 001416) using the following primers:

attPUP: CCTTGCGCTAATGCTCTGTTACAGG SEQ ID No. 1

attPDWN: CAGAGGCAGGGAGTGGGACAAAATTG SEQ ID No. 2

After amplification and purification of the resulting fragment, the attP site was cloned into the *SmaI* site of pSV40193 and the orientation of the attP site was determined by DNA sequence analysis (plasmid pSV40193attP). The gene encoding puromycin resistance (Puro) was isolated by digesting the plasmid pPUR (Clontech Laboratories, Inc. Palo Alto, CA) with *AgeI/BamHI* followed by filling in the overhangs with Klenow and subsequently cloned into the *Ascl* site downstream of the attP site of pSV40193attP generating the plasmid pSV40193attPsensePUR (Figure 4; SEQ ID NO:113)).

The plasmid pSV40193attPsensePUR was digested with *ScaI* and co-transfected with the plasmid pFK161 (SEQ ID NO: 118) into mouse LMtk- cells and platform artificial chromosomes were identified and isolated as described above. The process for generating this exemplary platform ACes containing multiple site-specific recombination sites is summarized in Figure 5. One platform ACes resulting from this experiment is designated B19-18. This platform ACes chromosome may subsequently be engineered to contain target gene expression nucleic acids using the lambda integrase mediated site-specific recombination system as described herein in Example 7 and 8.

EXAMPLE 4

Lambda integrase mediated site-specific recombination of a RFP expressing vector onto artificial chromosomes

In this example, a vector expressing the red fluorescent protein (RFP) was produced and recombined into the attP site residing on an artificial chromosome within LMTK- cells. This recombination is depicted in Figure 7.

A. Construction of expression vectors containing wildtype and mutant lambda integrase

Mutations at the glutamic acid at position 174 in the lambda integrase protein relaxes the requirement for the accessory protein IHF during recombination and DNA supercoiling *in vitro* (see, Miller *et al.* (1980) *Cell* 20:721-729; Lange-Gustafson *et al.* (1984) *J. Biol. Chem.* 259:12724-12732). Mutations at this site promote attP, attB intramolecular recombination in mammalian cells (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181).

To construct nucleic acid encoding the mutant, lambda integrase was PCR amplified from bacteriophage lambda DNA (cl857 *ind Sam* 7; New England Biolabs) using the following primers:

Lamint1 (SEQ ID No. 3)

TTCGAATTCATGGGAAGAAGGCGAAGTCATGAGCG)

Lamint2 (SEQ ID No. 4)

(TTCGAATTCTTATTTGATTTCAATTTGTCCCAC).

The resulting PCR product was digested with *EcoR* I and cloned into the

- 5 *EcoR* I site of pUC19. Lambda integrase was mutated at amino acid position 174 using QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the following oligos (generating a glutamic acid to arginine change at position 174):

LambdaINTE174R

- 10 (SEQ ID No. 6)

(CGCGCAGCAAAATCTAGAGTAAGGAGATCAAGACTTACGGCTGACG),

LamintR174rev (SEQ ID No. 7)

(CGTCAGCCGTAAGTCTTGATCTCCTTACTCTAGATTTTGCTGCGCG).

The resulting site directed mutant was confirmed by sequence analysis.

- 15 The wildtype and mutant lambda genes were cloned into the *EcoR* I site of pCX creating pCX-LamInt (SEQ ID NO: 127) and pCXLamIntR (Figure 8; SEQ ID NO: 112).

- 20 The plasmid pCX (SEQ ID No. 70) was derived from plasmid pCXeGFP (SEQ ID No. 71). Excision of the *EcoRI* fragment containing the eGFP marker generated pCX. To generate plasmid pCXLamINTR (SEQ ID NO: 112) an *EcoRI* fragment containing the lambda integrase E174R (SEQ ID No. 37) mutation was cloned into the *EcoRI* site of pCX, and to generate plasmid pCX-LamINT, an *EcoRI* fragment containing the wild-type lambda integrase was cloned into the *EcoRI* site of pCX.

- 25 **B. Construction of integration vector containing attB and DsRed**

The plasmid pDsRedN1 (Clontech Laboratories, Palo Alto, CA; SEQ ID No. 29) was digested with *Hpa* I and ligated to the following annealed oligos:

attB1 (SEQ ID No. 8)

(TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA)

attB2 (SEQ ID No. 9)

(TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA)

The resulting vector (pDsRedN1-attB) was confirmed by PCR and

5 sequence analysis.

C. Transfection into LMtk- cells

LM(tk-) cells containing the Prototype A ACes (L1-18; Chromos
Molecular Systems Inc., Burnaby, BC Canada) were co-transfected with
pDsRedN1 or pDsRedN1-attB and either pCXLamInt (SEQ ID NO: 127) or
10 pCXLamIntR (SEQ ID NO: 112) using Lipofectamine Plus Reagent
(LifeTechnologies, Gaithersburg, MD). The transfected cells were grown
in DMEM (LifeTechnologies, Gaithersburg, MD) with 10% FBS (CanSera)
and G418 (CalBiochem) at a concentration of 1 mg/ml.

D. Enrichment by cell sorting

15 The transfected cells were sorted using a FACs Vantage SE cell
sorter (Becton Dickenson) to enrich for cells expressing DsRed. The cells
were excited with a 488 nm Argon laser at 200 watts and cells
fluorescing in the 585/42 detection channel were collected. The sorted
cells were returned to growth medium for recovery and expansion. After
20 three successive enrichments for cells expressing DsRed, single cell
sorting into 96 well plates was performed using the same parameters.
Duplicate plates of the single cell clones were made for PCR analysis.

E. PCR analysis of single cell clones

25 Pools of cells from each row and column of the 96 well plate were
used for DNA isolation. DNA was prepared using a Wizard Genomic DNA
purification kit (Promega Inc, Madison, WI). Nested PCR analysis on the
DNA pools was performed to confirm the site-specific recombination
event using the following primer sets:

attPdwn2 (SEQ ID No. 10)
(TCTTCTCGGGCATAAGTCGGACACC)

CMVen (SEQ ID No. 11)
(CTCACGGGGATTCCAAGTCTCCAC)

5 followed by:

attPdwn (SEQ ID No. 12)
(CAGAGGCAGGGAGTGGGACAAAATTG)

CMVen2 (SEQ ID No. 13)
(CAACTCCGCCCCATTGACGCAAATG).

10 The resulting PCR reactions were analyzed by gel electrophoresis and the potential individual clones containing the site-specific recombination event were identified by combining the PCR results of all of the pooled rows and columns for each 96 well plate. The individual clones were then further analyzed by PCR using the following primers that flank the
15 recombination junction. L1for and F1rev flank the attR junction whereas REDfor and L2rev flank the attL junction (see Figure 7):

L1for (SEQ ID No. 14)
AGTATCGCCGAACGATTAGCTCTTCA

20 F1rev (SEQ ID No. 15)
GCCGATTTCGGCCTATTGGTTAAA

REDfor (SEQ ID No. 16)
CCGCCGACATCCCCGACTACAAGAA

L2rev (SEQ ID No. 17)
TTCCTTCGAAGGGGATCCGCCTACC.

25 **F. Sequence analysis of recombination junctions**

PCR products spanning the recombination junction were Topo-cloned into pcDNA3.1D/V5His (Invitrogen Inc., San Diego, CA) and then sequenced by cycle-sequencing. The clones were confirmed to have the correct *attR* and *attL* junctions by cycle sequencing.

30 **G. Fluorescent In Situ Hybridization (FISH)**

The cell lines containing the correct recombination junction sequence were further analyzed by fluorescent *in situ* hybridization (FISH)

by probing with the DsRed coding region labeled with biotin and visualizing with the Tyramide Signal Amplification system (TSA; NEN Life Science Products). The results indicate that the RFP sequence is present on the ACes.

5 **H. Southern analysis**

Genomic DNA was harvested from the cell lines containing an ACes with the correct recombinant event and digested with *EcoR* I. The digested DNAs were separated on a 0.7% agarose gel, transferred and fixed to a nylon membrane and probed with RFP coding sequences. The
10 result showed that there is an integrated copy of RFP coding sequence in each clone.

EXAMPLE 5

Delivery of a second gene encoding GFP onto the RFP platform ACes

15 **A. Construction of integration vector containing attB and GFP (pD2eGFPIresPuroattB).**

The plasmid pRESpuro2 (Clontech, Palo Alto, CA; SEQ ID NO: 88) was digested with *EcoRI* and *NotI* then ligated to the D2eGFP *EcoRI-NotI* fragment from pD2eGFP-N1 (Clontech, Palo Alto, CA) to create pD2eGFPIresPuro2. Subsequently, oligos encoding the attB site were
20 annealed and ligated into the *NruI* site of pD2eGFPIresPuro2 to create pD2eGFPIresPuroattB. The orientation of attB in the *NruI* site was determined by PCR.

B. Transfection of LMtk- cells

The LMtk- cells containing the RFP platform ACes produced in
25 Example 4, which has multiple attP sites, were co-transfected with pCXLamIntR and pD2eGFPIresPuroattB using LipofectAMINE PLUS reagent. Five μ g of each vector was placed into a tube containing 750 μ l of DMEM (Dulbecco's modified Eagles Medium). Twenty μ l of the Plus reagent was added to the DNA and incubated at room temperature for 15

minutes. A mixture of 30 μ l of lipofectamine and 750 μ l DMEM was added to the DNA mixture and incubated an additional 15 minutes at room temperature. The DNA mixture was then added dropwise to approximately 3 million cells attached to a 10cm dish in 5 mls of DMEM.

- 5 The cells were incubated 4 hours (37°C, 5% CO₂) with the DNA-lipid mixture, after which DMEM with 20% fetal bovine serum was added to the dishes to bring the culture medium to 10% fetal bovine serum. The dishes were incubated at 37°C with 5% CO₂.

Plasmid pD2eGFPIresPuroattB has a puromycin gene

- 10 transcriptionally linked to the GFP gene *via* an IRES element. Two days after the transfection the cells were placed in medium containing puromycin at 4 μ g/ml to select for cells containing the pD2eGFPIresPuroattB plasmid integrated into the genome. Twenty-three clones were isolated after 17 days of selection with puromycin. These
15 clones were expanded and then analyzed for the presence of the GFP gene on the ACes by 2-color (RFP/biotin & GFP/digoxigenin) TSA-FISH (NEN) according to the manufacturers protocol. Sixteen of the 23 clones produced a positive FISH signal on the ACes with a GFP probe.

EXAMPLE 6

- 20 **Delivery Of ACes Into human Mesenchymal Stem Cells (hMSC)**

A. Transfection

- Transfection conditions for the most efficient delivery of the ACes into hMSCs (Cambrex BioWhittaker Product Code PT-2501, lot# F0658, East Rutherford, New Jersey) were assayed using LipofectAMINE PLUS
25 and Superfect. One million prototype B ACes, which is a murine derived 60Mb ACes having primarily murine pericentric heterochromatin, and carrying a "payload" containing a hygromycin B selectable marker gene and a *lacZ* reporter gene (see , Telenius et al., 1999, Chrom. Res., 7:3-7; and Kereso et al., 1996, Chrom. Res., 4:226-239; each of which is

incorporated herein by reference in its entirety), were combined with 1-12 μ l of the transfection agent. In the case of LipofectAMINE PLUS, the PLUS reagent was combined with the ACes for 15 minutes followed by LipofectAMINE for a further 15 minutes. Superfect was complexed for 5 10 minutes at a ratio of 2 μ l Superfect per 1 million ACes. The ACes/transfection agent complex was then applied to 0.5 million recipient cells and the transfection was allowed to proceed according to the manufacturer's protocol. Percent transfected cells was determined on a FACS Vantage flow cytometer with argon laser tuned to 488 nm at 10 200mW and FITC fluorescence collected through a standard FITC 530/30 nm band pass filter. After 24 hours, IdUrd labeled ACes were delivered to human MSCs in the range of 30-50%, varying with transfection agent and dose. ACes delivery curves were generated from data collected in experiments that varied the dose of the transfection reagents. Dose 15 response curves of Superfect and LipofectAMINE PLUS, showing delivery of ACes into recipient hMSCs cells, were prepared, measured by transfer of IdUrd labeled ACes and detected by flow cytometry. Superfect shows maximum delivery in the range of 30-50% at doses greater than 2 μ l per million ACes. LipofectAMINE PLUS has a 42-48% delivery peak around 20 5-8 μ l per million ACes. These dose curves were then correlated with toxicity data to determine the transfection conditions that will allow for highest potential transfection efficiency. Toxicity was determined by a modified plating efficiency assay (de Jong et al., 2001, Chrom. Research, 9:475-485). The population's normalized plating efficiency (at maximum 25 % delivery doses) was in the range of 0.2 - 0.4 for Superfect and 0.5 - 0.6 with LipofectAMINE PLUS.

Due to the transfected population consisting of mixed cell types, flow cytometry allowed for the assessment of ACes delivery into each sub-population and the purification of the target population. Flow profiles

showing forward scatter (cell size) and side scatter (internal cell granularity) revealed three distinct hMSC populations that were gated into three regions: R3 (small cell region), R4 (medium cell region), R5 (large cell region). Transfection conditions were further optimized by re-analyzing delivery curves and assessing the differences in delivery to each sub-population. Dose response curves of Superfect and LipofectAMINE were prepared showing % delivery to each sub-population represented by the gating on basis of cell size and granularity properties of the mixed population. Three distinct hMSC populations were gated and % delivery dose curves generated. Using Superfect and LipofectAMINE PLUS the overall % delivery increased with cell size (80-90% delivery in large cells). LipofectAMINE PLUS at high doses (8-12 μ l per 1 million ACes) shows an increase in the overall proportion of chromosome transfer to the small population (10-20%). This suggests an advantage to using this transfection agent if the small-undifferentiated cell population is the desired target host cell.

B. Expression from Genes on ACes IN hMSCs

Following the delivery screening process conducted in section (A) above, the most promising results were subjected to further analyses to monitor expression and verify the presence of structurally intact ACes. The transfection conditions employed for these experiments were exactly the same as those that had been used during the screening process. Short-term expression was monitored by transfecting hMSCs with ACes containing a RFP gene (red fluorescent protein) set forth in Example 2C as "D11C4". The unselected population was harvested at 72-96 hours post transfection and % positive fluorescent cells measured by flow cytometry. RFP expression was in the range of 1-20%.

Long term-gene expression was assayed by selecting for hygromycin B resistant cells over a period of 7-10 days. Cytogenetic

analysis was done to detect presence of intact *ACes* by Fluorescent *In Situ* hybridization (FISH), where metaphase chromosomes were hybridized to a mouse major satellite-DNA probe (targeting murine pericentric heterochromatin) and a lambda probe (hybridizing to the *lacZ* gene). The human mesenchymal transfected culture could not undergo standard sub-cloning as diffuse colonies form with limited doublings available for expansion. Cytogenetic analysis was performed on the entire population, sampling over a period of 3-10 days post-transfection. The hygromycin resistant population was then blocked in mitosis with colchicine and analyzed for presence of intact *ACes* by FISH. Preliminary FISH results show approximately 2-8% of the hMSC-transfected population had an intact *ACes*. This compared to rat skeletal muscle myoblast clones, which were in the range of 60-95%. To increase the % of intact *ACes* in the hMSC-transfected population an enrichment step can be utilized as described in Example 2C.

C. Differentiation of The hMSCs

In initial experiments where transfected hMSCs cells have been induced to differentiate into adipose or osteocytes, the results indicate that the transfected cells appear to be differentiating at a rate comparable to the untransfected controls and the cultures are lineage specific as tested by microscopic examination, FISH, Oil Red O staining (adipocyte assay), and calcium secretion (osteocyte assay).

Accordingly, these results indicate that the artificial chromosomes (*ACes*) provided herein can be successfully transferred into hMSC target cells. Targeting MSCs (such as hMSCs) permits gene transfer into cells in an undifferentiated state where the cells are easier to expand and purify. The genetically modified cells can then be differentiated *in vitro* or injected into a site *in vivo* where the microenvironment will induce transformation into specific cell lineages.

EXAMPLE 7

Delivery of a Promoterless Marker Gene to a Platform ACes

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

5 **A. Construction of Targeting Vectors.**

The base vector p18attBZeo (3166bp; SEQ ID NO: 114) was constructed by ligating the 1067bp *HindIII-SspI* fragment containing attBZeo, obtained from pLITattBZeo (SEQ ID NO:91), into pUC18 (SEQ ID NO: 122) digested with *HindIII* and *SspI*.

10 1. p18attBZEO-eGFP (6119bp; SEQ ID NO: 126) was constructed by inserting the 2977bp *SpeI-HindIII* fragment from pCXeGFP (SEQ ID NO:71; Okabe, *et al.* (1997) *FEBS Lett* 407:313-319) containing the eGFP gene into p18attBZeo (SEQ ID NO: 114) digested with *HindIII* and *XbaI*.

15 2. p18attBZEO-5'6XHS4eGFP (Figure 10; 7631bp; SEQ ID NO: 116) was constructed by ligating the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123), which contains the eGFP gene under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal, into the *HindIII* site of p18attBZeo (SEQ ID NO: 20 114).

 3. p18attBZEO-3'6XHS4eGFP (Figure 11; 7600bp; SEQ ID NO: 115) was created by removing the 5'6XHS4 element from p18attBZeo-(6XHS4)2eGFP (SEQ ID NO: 110). p18attBZeo-(6XHS4)2eGFP was digested with *EcoRV* and *SpeI*, treated with Klenow and religated to form 25 p18attBZeo3'6XHS4eGFP (SEQ ID NO: 115).

 4. p18attBZEO-(6XHS4)2eGFP (Figure 12; 9080bp; SEQ ID NO: 110) was created in two steps. First, the *EcoRI-SpeI* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123), which contains 6 copies of the HS4 core element, was ligated into p18attBZeo (SEQ ID NO: 114)

digested with *EcoRI* and *XbaI* to create p18attBZeo6XHS4 (4615bp; SEQ ID NO: 117). Next, p18attBZeo6XHS4 was digested with *HindIII* and ligated to the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 which contains the eGFP gene under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal.

Table 2

Targeting plasmid	No. zeocin resistant clones	No. clones with expected PCR product size	No. clones with correct sequence at recombination junction
p18attBZEO-eGFP	12	12	NT*
p18attBZEO-5'6XHS4eGFP	11	11	NT
p18attBZEO-3'6XHS4eGFP	11	11	NT
p18attBZEO-(6XHS4)2eGFP	9	9	4/4

*NT = not tested

B. Transfection and Selection with Drug.

The mouse cell line containing the 2nd generation platform ACE, B19-38 (constructed as set forth in Example 3), was plated onto four 10cm dishes at approximately 5 million cells per dish. The cells were incubated overnight in DMEM with 10% fetal calf serum at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each of the 4 vectors listed in Example 7.A. above and 5µg of pCXLamIntR (SEQ ID NO: 112), for a total of 10µg per 10cm dish. Lipofectamine Plus reagent was used to transfect the cells according to the manufacturers protocol. Two days post-transfection zeocin was added to the medium at 500µg/ml. The cells were maintained in selective medium until colonies formed. The colonies were then ring-cloned (see, e.g., McFarland, 2000, Methods Cell Sci, Mar;22(1):63-66).

C. Analysis of Clones (PCR, SEQUENCING).

Genomic DNA was isolated from each of the candidate clones with the Wizard kit (Promega) and following the manufacturers protocol. The following primer set was used to analyze the genomic DNA isolated from the zeocin resistant clones: 5PacSV40 –

- 5 CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO:76);
Antisense Zeo - TGAACAGGGTCACGTCGTCC (SEQ ID NO:77). PCR amplification with the above primers and genomic DNA from the site-specific integration of any of the 4 zeocin vectors would result in a 673bp PCR product.
- 10 As set forth in Table 2, of the 4 zeocin resistant candidate clones thusfar analyzed by PCR, all 4 exhibit the correct sequence for a site-specific integration event.

EXAMPLE 8

Integration of a PCR product by site-specific recombination.

- 15 In this example a gene is integrated onto the platform ACes by site-specific recombination without cloning said gene into a vector.

A. PCR PRIMER DESIGN.

- 20 PCR primers are designed to contain an attB site at the 5' end of one of the primers in the primer set. The remaining primers, which could be one or more than one primer, do not contain an attB site, but are complementary to sequences flanking the gene or genes of interest and any associated regulatory sequences. In first example, 2 primers (one containing an attB site) are used to amplify a selective gene such as puromycin.

- 25 In a second example as shown in Figure 13, the primer set includes primers 1 & 2 that amplify the GFP gene without amplification of an upstream promoter. Primer 1 contains the attB site at the 5' end of the oligo. Primers 3 & 4 are designed to amplify the IRES-blasticidin DNA sequences from the vector pIRESblasticidin. The 5' end of primer 3

contains sequences complementary to the 5' end of primer 2 such that annealing can occur between 5' ends of the two primers.

B. PCR REACTION AND SUBSEQUENT LIGATION TO CREATE CIRCULAR MOLECULES FROM THE PCR PRODUCT

- 5 In the first example set forth above in Section A, the two PCR primers are combined with a puromycin DNA template such as pPUR (Clontech), a heat stable DNA polymerase and appropriate conditions for DNA amplification. The resulting PCR product (attB-Puromycin) is then then purified and self-ligated to form a circular molecule.
- 10 In the second example set forth above in Section A, amplification of the GFP gene and IRES-blasticidin sequences is accomplished by combining primers 1 & 2 with DNA template pD2eGFP and primers 3 & 4 with template pIRESblasticidin under appropriate conditions to amplify the desired template. After initial amplification of the two products (attB-GFP
- 15 & IRES-blasticidin) in separate reactions, a second round of amplification using both of the PCR products from the first round of amplification together with primers 1 and 4 amplifies the fusion product attB-GFP-IRES-blasticidin (Figure 13). This technique of using complementary sequences in primer design to create a fusion product is employed in *Saccharomyces*
- 20 *cerevisiae* for allele replacement (Erdeniz *et al* (1997) *Gen Res* 7:1174-1183). The amplified product is then purified from the PCR reaction mixture by standard methods and ligated to form a circular molecule.

C. INTRODUCTION OF PCR PRODUCT ONTO THE ACes USING A RECOMBINASE

- 25 The circular PCR product is then be introduced to the platform ACes using the bacteriophage lambda integrase E174R. The introduction can be performed *in vivo* by transfecting the pCXLamIntR (SEQ ID NO: 112) vector encoding the lambda integrase mutant E174R together with the circularized PCR product into a cell line containing the platform ACE.

D. SELECTION FOR MARKER GENE

The marker gene (in this case either puromycin, blasticidin or GFP) is used to enrich the population for cells containing the proper integration event. A proper integration event in the second example (Figure 14) juxtaposes a promoter residing on the platform *ACes* 5' to the attB-GFP-IRES-Blasticidin PCR product, allowing for transcription of both GFP and blasticidin. If enrichment is done by drug selection, blasticidin is added to the medium on the transfected cells 24-48 hours post-transfection. Selection is maintained until colonies are formed on the plates. If enrichment is done by cell sorting, cells are sorted 2-4 days post-transfection to enrich for cells expressing the fluorescent marker (GFP in this case).

E. ANALYSIS OF CLONES

Clonal isolates are analyzed by PCR, FISH and sequence analysis to confirm proper integration events.

EXAMPLE 9

Construction of a human platform *ACes* "ACE 0.1"

A. CONSTRUCTION OF THE TARGETING VECTOR pPACrDNA

Genome Systems (IncyteGenomics) was supplied with the primers 5'HETS (GGGCCGAAACGATCTCAACCTATT; SEQ ID NO:78), and 3'HETS (CGCAGCGGCCCTCCTACTC; SEQ ID NO:79), which were used to amplify a 538bp PCR product homologous to nt 9680-10218 of the human rDNA sequences (GenBank Accession No. U13369) and used as a probe to screen a human genomic P1AC (P1 Artificial Chromosome) library constructed in the vector pCYPAC2 (Ioannou *et al.* (1994) *Nat. Genet.* 6(1): 84-89). Genome Systems clone #18720 was isolated in this screen and contains three repeats of human rDNA as assessed by restriction analysis. GS clone #18720, was digested with PmeI, a restriction enzyme unique to a single repeat of the human rDNA (45Kbp),

and then religated to form pPACrDNA (Figure 15). The insert in pPACrDNA was analyzed by restriction digests and sequence analysis of the 5' and 3' termini. The pPACrDNA, rDNA sequences are homologous to Genbank Accession #U13369, containing an insert of about 45 kB

5 comprising a single repeat beginning from the end of one repeat at ~33980 (relative to the Genbank sequence) through the beginning of the next repeat up to approximately 35120 (the repeat offset from that listed in the GenBank file). Thus, the rDNA sequence is just over 1 copy of the repeat extending from 33980 (+/-10bp) to the end of the first repeat

10 (43Kbp) and continuing into the second repeat to bp 35120 (+/-10bp).

B. TRANSFECTION AND ACes FORMATION.

Five hundred thousand MSU1.1 cells (Morgan et al., 1991, Exp. Cell Res., Nov;197(1):125-136; provided by Dr. Justin McCormick at Michigan State University) were plated per 6cm plate (3 plates total) and

15 allowed to grow overnight. The cells were 70-80% confluent the following day. One plate was transfected with 15µg pPACrDNA (linearized with *Pme* I) and 2µg pSV40attPsensePuro (linearized with *Sca* I; see Example 3). The remaining plates were controls and were transfected with either 20µg pBS (Stratagene) or 20µg

20 pSV40attBsensePuro (linearized with *Sca* I). All three plates were transfected using a CaPO₄ protocol.

C. SELECTION OF PUROMYCIN RESISTANT COLONIES

One day post-transfection the cells were "glycerol shocked" by the addition of PBS medium containing 10% glycerol for 30 seconds.

25 Subsequently, the glycerol was removed and replaced with fresh DMEM. Four days post-transfection selective medium was added. Selective medium contains 1µg/ml puromycin. The transfection plates were maintained at 37°C with 5% CO₂ in selective medium for 2 weeks at which point colonies could be seen on the plate transfected with

pPACrDNA and pSV40attPsensePuro. The colonies were ring-cloned from the plate on day 17 post-selection and expanded in selective medium for analysis. Only two colonies (M2-2d & M2-2b) were able to proliferate in the selective medium after cloning. No colonies were seen on the control plates after 37 days in selective medium.

D. ANALYSIS OF CLONES

FISH analysis was performed on the candidate clones to detect ACes formation. Metaphase spreads from the candidate clones were probed in multiple probe combinations. In one experiment, the probes used were biotin-labeled human alphoid DNA (pPACrDNA) and digoxigenin-labeled mouse major DNA (pFK161) as a negative control. Candidate M2-2d was single cell subcloned by flow sorting and the candidate subclones were reanalyzed by FISH. Subclone 1B1 of M2-2d was determined to be a platform ACes and is also designated human Platform ACE 0.1.

EXAMPLE 10

Site-specific integration of a marker gene onto a human platform ACE 0.1

The promoterless delivery method was used to deliver a promoterless blasticidin marker gene onto the human platform ACes with excellent results. The human ACes platform with a promoterless blasticidin marker gene resulted in 21 of 38 blasticidin resistant clones displaying a PCR product of the expected size from the population co-transfected with pLIT38attBBSRpolyA10 and pCXLamIntR (Figure 8; SEQ ID NOs. 111 and 112). Whereas, the population transfected with pBlueScript resulted in 0 blasticidin resistant colonies.

A. CONSTRUCTION OF pLIT38attB-BSRpolyA10 & pLIT38attB-BSRpolyA2.

The vector pLITMUS 38 (New England Biolabs; U.S. Patent No. 5,691,140; SEQ ID NO: 119) was digested with *EcoRV* and ligated to

two annealed oligomers, which form an attB site (attB1 5'-TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA-3' (SEQ ID NO:8); attB2 5'- TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA-3'; SEQ ID NO:9). This ligation reaction resulted in the vector pLIT38attB (SEQ ID NO: 120).

- 5 The blasticidin resistance gene and SV40 polyA site were PCR amplified with primers: 5BSD (ACCATGAAAACATTTAACATTTCTCAACA; SEQ ID NO:80) and SV40polyA (TTTATTTGTGAAATTTGTGATGCTATTGC; SEQ ID NO:81) using pPAC4 (Frengen, E., *et al.* (2000) Genomics 68 (2), 118-126; GenBank Accession No. U75992) as template. The blasticidin-SV40polyA PCR product was then ligated into pLIT38attB at the *Bam*HI site, which was Klenow treated following digestion with *Bam*HI. pLIT38attB-BSDpolyA10 (SEQ ID NO: 111) and pLIT38attB-BSDpolyA2 (SEQ ID NO: 121) are the two resulting orientations of the PCR product ligated into the vector.

15 **B. TRANSFECTION OF MSU1.1 CELLS CONTAINING HUMAN PLATFORM ACE 0.1.**

- MSU1.1 cells containing human platform ACE 0.1 (see Example 9) were expanded and plated to five 10cm dishes with 1.3×10^6 cells per dish. The cells were incubated overnight in DMEM with 10% fetal bovine serum, at 37°C and 5% CO₂. The following day the cells were transfected with 5µg of each plasmid as set forth in Table 3, for a total of 10µg of DNA per plate of cells transfected (see Table 3) using ExGen 500 *in vitro* transfection reagent (MBI fermentas, cat. no. R0511). The transfection was performed according to the manufacturers protocol.
- 20 Cells were incubated at 37°C with 5% CO₂ in DMEM with 10% fetal bovine serum following the transfection.
- 25

Table 3

Plate #	Plasmid 1	Plasmid 2	No. Bsd ^R Colonies
1	pBS	None	0
2	pCXLamInt	pLIT38attB-BSRpolyA10	16
3	pCXLamIntR	pLIT38attB-BSRpolyA10	16
4	pCXLamInt	pLIT38attB-BSRpolyA2	28
5	pCXLamIntR	pLIT38attB-BSRpolyA2	36

10 **C. SELECTION OF BLASTICIDIN RESISTANT CLONES.**

Three days following the transfection the cells were split from a 10 cm dish to two 15cm dishes. The cells were maintained in DMEM with 10% fetal bovine serum for 4 days in the 15 cm dishes. Seven days post-transfection blasticidin was introduced into the medium. Stably transfected cells were selected with 1 μ g/ml blasticidin. The number of colonies formed on each plate is listed in Table 3. These colonies were ring-cloned and expanded for PCR analysis. Upon expansion in blasticidin containing medium some clones failed to live and therefore do not have corresponding PCR data.

20 **D. PCR ANALYSIS**

Thirty-eight of the 40 clones from plate 3 grew after ring-cloning. Genomic DNA was isolated from these clones with the Promega Wizard Genomic cDNA purification kit, digested with *Eco*RI and used as template in a PCR reaction with the following primers: 3BSP – TTAATTTCTGGG TATATTTGAGTGGA (SEQ ID NO:82); 5PacSV40 – CTGTTAATTAAGTGTGGAA TGTGTGTCTAGTTAGGGTG (SEQ ID NO:76). The PCR conditions were as follows. 100ng of genomic DNA was

amplified with 0.5 μ l Herculanase polymerase (Stratagene) in a 50 μ l reaction that contained 12.5pmole of each primer, 2.5mM of each dNTP, and 1X Herculanase buffer (Stratagene). The reactions were placed in a PerkinElmer thermocycler programmed as follows: Initial denaturation at 95°C for 10 minutes; 35 cycles of 94°C for 1 minute, 53°C for 1 minute, 72°C for 1 minute, and 72°C for 1 minute; Final extension for 10 minutes at 72°C; and 4°C hold. If pLIT38attB-BSRpolyA10 integrates onto the human platform ACE 0.1 correctly, PCR amplification with the above primers should yield an 804bp product. Twenty-one of the 38 clones from plate 3 produced a PCR product of the expected 804bp size.

EXAMPLE 11

Delivery of a Vector comprising a Promoterless Marker Gene and a gene encoding a therapeutic product to a Platform ACes

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

A. CONSTRUCTION OF DELIVERY VECTORS

1. Erythropoietin cDNA vector, p18EPOcDNA.

The erythropoietin cDNA was PCR amplified from a human cDNA library (E. Perkins *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96(5): 2204-2209) using the following primers: EPO5XBA - TATCTAGAATGGGGGTGC ACGAATGTCCTGCC (SEQ ID NO: 83); EPO3BSI - TACGTACGTCATC TGTCCCCTGTCCTGCAGGC (SEQ ID NO: 84). The cDNA was amplified through two successive rounds of PCR using the following conditions: heat denaturation at 95°C for 3 minutes; 35 cycles of a 30 second denaturation (95°C), 30 seconds of annealing (60°C), and 1 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. BIO-X-ACT (BIOLINE) was used to amplify the erythropoietin cDNA from 2.5ng of the human cDNA library in the first round of amplification. Five μ l of the first amplification product was used

as template for the second round of amplification. Two PCR products were produced from the second amplification with Taq polymerase (Eppendorf), each product was cloned into pCR2.1-Topo (Invitrogen) and sequenced. The larger PCR product contained the expected cDNA
5 sequence for erythropoietin. The erythropoietin cDNA was moved from pTopoEPO into p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110). pTopoEPO was digested with BsiWI and XbaI to release a 588 bp EPO cDNA. BsrGI and BsiWI create compatible ends. The eGFP gene was removed from p18attBZeo(6XHS4)2eGFP by digestion with BsiWI and XbaI, the 8.3 Kbp
10 vector backbone was gel purified and ligated to the 588 bp EPO cDNA to create p18EPOcDNA (SEQ ID NO: 124).

2. Genomic erythropoietin vector, p18genEPO.

The erythropoietin genomic clone was PCR amplified from a human genomic library (Clontech) using the following primers: GENEPO3BSI -
15 CGTACGTCATCTGTCCCCT GTCCTGCA (SEQ ID NO: 85); GENEPO 5XBA -TCTAGAATGGGGGT GCACGGTGAGTACT (SEQ ID NO: 86). The reaction conditions for the amplification were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to
20 50°C), and 3 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds annealing (50°C), and 3 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. The erythropoietin genomic PCR product (2147 bp) was gel purified and cloned into pCR2.1Topo to create pTopogenEPO. Sequence
25 analysis revealed 2bp substitutions and insertions in the intronic sequences of the genomic clone of erythropoietin. A partial digest with XbaI and complete digest with BsiWI excised the erythropoietin genomic insert from pTopogenEPO. The resulting 2158 bp genomic erythropoietin fragment was ligated into the 8.3 Kbp fragment resulting from the

digestion of p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110) with XbaI and BsrGI to create p18genEPO (SEQ ID NO: 125).

B. TRANSFECTION AND SELECTION WITH DRUG

5 The erythropoietin genomic and cDNA genes were each moved onto the platform *ACes* B19-38 (constructed as set forth in Example 3) by co-transfecting with pCXLamIntR. Control transfections were also performed using pCXLamInt (SEQ ID NO: 127) together with either p18EPOcDNA (SEQ ID NO: 124) or p18genEPO (SEQ ID NO: 125). Lipofectamine Plus was used to transfect the DNA's into B19-38 cells
10 according to the manufacturer's protocol. The cells were placed in selective medium (DMEM with 10% FBS and Zeocin @ 500ug/ml) 48 hours post-transfection and maintained in selective medium for 13 days. Clones were isolated 15 days post-transfection.

C. ANALYSIS OF CLONES (ELISA, PCR)

15 **1. ELISA Assays**

Thirty clones were tested for erythropoietin production by an ELISA assay using a monoclonal anti-human erythropoietin antibody (R&D Systems, Catalogue # MAB287), a polyclonal anti-human erythropoietin antibody (R & D Systems, Catalogue # AB-286-NA) and alkaline
20 phosphatase conjugated goat-anti-rabbit IgG (heavy and light chains) (Jackson ImmunoResearch Laboratories, Inc., Catalogue # 111-055-144). The negative control was a Zeocin resistant clone isolated from B19-38 cells transfected with p18attBZeo(6XHS4) (SEQ ID NO: 117; no insert control vector) and pCXLamIntR (SEQ ID NO: 112). The preliminary
25 ELISA assay was executed as follows: 1) Nunc-Immuno Plates (MaxiSorb 96-well, Catalogue # 439454) were coated with 75µl of a 1/200 dilution (in Phosphate buffered Saline, pH 7.4 (PBS), Sigma Catalogue # P-3813) of monoclonal anti-human erythropoietin antibody overnight at 4°C. 2) The following day the plates were washed 3 times with 300µl PBS

containing 0.15% Tween 20 (Sigma, Catalogue # P-9416). 3) The plates were then blocked with 300 μ l of 1% Bovine Serum Albumin (BSA; Sigma Catalogue # A-7030) in PBS for 1 hour at 37°C. 4) Repeat the washes as in step 2. 5) The clonal supernatants (75 μ l per clone per well of 96-well plate) were then added to the plate and incubated for 1 hour at 37°C. The clonal supernatant analyzed in the ELISA assay had been maintained on the cells 7 days prior to analysis. 6) Repeat the washes of step 2. 7) Add 75 μ l of polyclonal anti-human erythropoietin antibody (1/250 dilution in dilution buffer (0.5% BSA, 0.01% Tween 20, 1X PBS, pH 7.4) and incubate 1 hour at 37°C. 8) Repeat washes of step 2. 9) Add 75 μ l of goat anti-rabbit conjugated alkaline phosphatase diluted 1/4000 in dilution buffer and incubate 1 hour at 37°C. 10) Repeat washes of step 2. 11) Add 75 μ l substrate, p-nitrophenyl phosphate (Sigma N2640), diluted to 1mg/ml in substrate buffer (0.1 Ethanolamine-HCl (Sigma, Catalogue # E-6133), 5mM MgCl₂ (Sigma, Catalogue # M-2393), pH 9.8). Incubate the plates in the dark for 1 hour at room temperature (22°C). 12) Read the absorption at 405nm (reference wavelength 495nm) on an Universal Microplate Reader (Bio-Tek Instruments, Inc., model # ELX800 UV). The erythropoietin standard curve was derived from readings of diluted human recombinant Erythropoietin (Roche, catalogue # 1-120-166; dilution range 125 - 7.8mUnits/ml). From this preliminary assay the 21 clones displaying the highest expression of erythropoietin were analyzed a second time in the same manner using medium supernatants that had been on the clones for 24 hours and a 1:3 dilution thereof.

25 2. PCR Analysis

Genomic DNA was isolated from the 21 clones with the best expression (as assessed by the initial ELISA assay above) as well as the B19-38 cell line and used for PCR analysis. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) according to the

manufacturers protocol. Amplification was performed on 100ng of genomic DNA as template with MasterTaq DNA Polymerase (Eppendorf) and the primer set 5PacSV40 – CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO: 76) and Antisense Zeo -

- 5 TGAACAGGGTCACGTCGTCC (SEQ ID NO: 77). The amplification conditions were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to 50°C), and 1 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds
10 annealing (50°C), and 1 minute extension (72°C); the last cycle is followed by a 10 minute extension at 72°C. PCR products were size separated by gel electrophoresis. Of the 21 clones analyzed 19 produced a PCR product of 650 bp as expected for a site-specific integration event. All nineteen clones were the result of transformations with p19EPOcDNA
15 (5) or p18genEPO (14) and pCXLamIntR (i.e. mutant integrase). The remaining two clones, both of which were the result of transformation with p18genEPO (SEQ ID NO: 125) and pCXLamInt (i.e. wildtype integrase; SEQ ID NO: 127), produced a 400 bp PCR product.

EXAMPLE 12

20 Preparation of a Transformation Vector Useful for the Induction of Plant Artificial Chromosome Formation

- Plant artificial chromosomes (PACs) can be generated by introducing nucleic acid, such as DNA, which can include a targeting DNA, for example rDNA or lambda DNA, into a plant cell, allowing the cell
25 to grow, and then identifying from among the resulting cells those that include a chromosome with a structure that is distinct from that of any chromosome that existed in the cell prior to introduction of the nucleic acid. The structure of a PAC reflects amplification of chromosomal DNA, for example, segmented, repeat region-containing and heterochromatic

structures. It is also possible to select cells that contain structures that are precursors to PACs, for example, chromosomes containing more than one centromere and/or fragments thereof, and culture and/or manipulate them to ultimately generate a PAC within the cell.

5 In the method of generating PACs, the nucleic acid can be introduced into a variety of plant cells. The nucleic acid can include targeting DNA and/or a plant expressable DNA encoding one or multiple selectable markers (*e.g.*, DNA encoding bialophos (bar) resistance) or scorable markers (*e.g.*, DNA encoding GFP). Examples of targeting DNA
10 include, but are not limited to, *N. tabacum* rDNA intergenic spacer sequence (IGS) and *Arabidopsis* rDNA such as the 18S, 5.8S, 26S rDNA and/or the intergenic spacer sequence. The DNA can be introduced using a variety of methods, including, but not limited to *Agrobacterium*-mediated methods, PEG-mediated DNA uptake and electroporation using,
15 for example, standard procedures according to Hartmann *et al* [(1998) *Plant Molecular Biology* 36:741]. The cell into which such DNA is introduced can be grown under selective conditions and can initially be grown under non-selective conditions and then transferred to selective media. The cells or protoplasts can be placed on plates containing a
20 selection agent to grow, for example, individual calli. Resistant calli can be scored for scorable marker expression. Metaphase spreads of resistant cultures can be prepared, and the metaphase chromosomes examined by FISH analysis using specific probes in order to detect amplification of regions of the chromosomes. Cells that have artificial chromosomes with
25 functioning centromeres or artificial chromosomal intermediate structures, including, but not limited to, dicentric chromosomes, formerly dicentric chromosomes, minichromosomes, heterochromatin structures (*e.g.* sausage chromosomes), and stable self-replicating artificial chromosomal intermediates as described herein, are

identified and cultured. In particular, the cells containing self-replicating artificial chromosomes are identified.

The DNA introduced into a plant cell for the generation of PACs can be in any form, including in the form of a vector. An exemplary
5 vector for use in methods of generating PACs can be prepared as follows.

For the production of artificial chromosomes, plant transformation vectors, as exemplified by pAgIIa and pAgIIb, containing a selectable marker, a targeting sequence, and a scorable marker were constructed using procedures well known in the art to combine the various fragments.
10 The vectors can be prepared using vector pAg1 as a base vector and inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase under the control of the nopaline synthase (NOS) promoter fragment and flanked at the 3' end by the NOS terminator fragment, a fragment of mouse satellite DNA and an *N. tabacum* rDNA intergenic
15 spacer sequence (IGS). In constructing plant transformation vectors, vector pAg2 can also be used as the base vector.

1. Construction of pAg1

Vector pAg1 (SEQ. ID. NO: 89) is a derivative of the CAMBIA vector named pCambia 3300 (Center for the Application of Molecular
20 Biology to International Agriculture, i.e., CAMBIA, Canberra, Australia; www.cambia.org), which is a modified version of vector pCambia 1300 to which has been added DNA from the bar gene conferring resistance to phosphinothricin. The nucleotide sequence of pCambia 3300 is provided in SEQ. ID. NO: 90. pCambia 3300 also contains a lacZ alpha sequence
25 containing a polylinker region.

pAg1 was constructed by inserting two new functional DNA fragments into the polylinker of pCambia 3300: one sequence containing an attB site and a promoterless zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal sequence, and a second

sequence containing DNA from the hygromycin resistance gene (hygromycin phosphotransferase) conferring resistance to hygromycin for selection in plants. Although the zeomycin-SV40 polyA signal fusion is not expected to function in plant cells, it can be activated in mammalian
5 cells by insertion of a functional promoter element into the attB site by site-specific recombination catalyzed by the Lambda att integrase. Thus, the inclusion of the attB-zeomycin sequences allows for evaluation of functionality of plant artificial chromosomes in mammalian cells by activation of the zeomycin resistance-encoding DNA, and provides an att
10 site for further insertion of new DNA sequences into plant artificial chromosomes formed as a result of using pAg1 for plant transformation. The second functional DNA fragment allows for selection of plant cells with hygromycin. Thus, pAg1 contains DNA from the bar gene conferring resistance to phosphinothricin, DNA from the hygromycin resistance gene,
15 both resistance-encoding DNAs under the control of a separate cauliflower mosaic virus (CaMV) 35S promoter, and the attB-promoterless zeomycin resistance-encoding DNA.

pAg1 is a binary vector containing *Agrobacterium* right and left T-DNA border sequences for use in *Agrobacterium*-mediated transformation
20 of plant cells or protoplasts with the DNA located between the border sequences. pAg1 also contains the pBR322 Ori for replication in *E.coli*. pAg1 was constructed by ligating *Hind*III/*Pst*I-digested p3300attBZeo with *Hind*III/*Pst*I-digested pBSCaMV35SHyg as follows.

a. Generation of p3300attBZeo

25 Plasmid pCambia 3300 was digested with *Pst*I/*Ec*136 II and ligated with *Pst*I/*Stu*I-digested pLITattBZeo (the nucleotide sequence of pLITattBZeo is provided in SEQ. ID. NO: 91), which contains DNA encoding the zeocin resistance gene and an attB Integrase recognition sequence, to generate p3300attBZeo which contains an attB site, a
30 promoterless

zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal, and a reconstructed *Pst*I site.

b. Generation of pBSCaMV35SHyg

A DNA fragment containing DNA encoding hygromycin

5 phosphotransferase flanked by the CaMV 35S promoter and the CaMV 35S polyA signal sequence was obtained by PCR amplification of plasmid pCambia 1302 (GenBank Accession No. AF234298 and SEQ. ID. NO: 92). The primers used in the amplification reaction were as follows: CaMV35SpolyA:

10 5'-CTGAATTAACGCCGAATTAATTCGGGGGATCTG-3' SEQ. ID. NO: 93
CaMV35Spr:

5'-CTAGAGCAGCTTGCCAACATGGTGGAGCA-3' SEQ. ID. NO: 94

The 2100-bp PCR fragment was ligated with *Eco*RV-digested pBluescript II SK+ (Stratagene, La Jolla, CA, U.S.A.) to generate pBSCaMV35SHyg.

15 **c. Generation of pAg1**

To generate pAg1, pBSCaMV35SHyg was digested with *Hind*III/*Pst*I and ligated with *Hind*III/*Pst*I-digested p3300attBZeo. Thus, pAg1 contains the pCambia 3300 backbone with DNA conferring resistance to phosphinothricin and hygromycin under the control of
20 separate CaMV 35S promoters, an attB-promoterless zeomycin resistance-encoding DNA recombination cassette and unique sites for adding additional markers, *e.g.*, DNA encoding GFP. The attB site can be used as described herein for the addition of new DNA sequences to plant artificial chromosomes, including PACs formed as a result of using the
25 pAg1 vector, or derivatives thereof, in the production of PACs. The attB site provides a convenient site for recombinase-mediated insertion of DNAs containing a homologous att site.

2. pAG2

The vector pAg2 (SEQ. ID. NO: 95) is a derivative of vector pAg1 formed by adding DNA encoding a green fluorescent protein (GFP), under the control of a NOS promoter and flanked at the 3' end by a NOS polyA signal, to pAg1. pAg2 was constructed as follows. A DNA fragment
5 containing the NOS promoter was obtained by digestion of pGEM-T-NOS, or pGEMEasyNOS (SEQ. ID. NO: 96), containing the NOS promoter in the cloning vector pGEM-T-Easy (Promega Biotech, Madison, WI, U.S.A.), with *Xba*I/*Nco*I and was ligated to a *Xba*I/*Nco*I fragment of pCambia 1302 containing DNA encoding GFP (without the CaMV 35S promoter) to
10 generate p1302NOS (SEQ. ID. NO: 97) containing GFP-encoding DNA in operable association with the NOS promoter. Plasmid p1302NOS was digested with *Sma*I/*Bs*WI to yield a fragment containing the NOS promoter and GFP-encoding DNA. The fragment was ligated with *Pme*I/*Bs*WI-digested pAg1 to generate pAg2. Thus, pAg2 contains DNA
15 from the bar gene conferring resistance to phosphinothricin, DNA conferring resistance to hygromycin, both resistance-encoding DNAs under the control of a cauliflower mosaic virus 35S promoter, DNA encoding kanamycin resistance, a GFP gene under the control of a NOS promoter and the attB-zeomycin resistance-encoding DNA. One of skill in
20 the art will appreciate that other fragments can be used to generate the pAg1 and pAg2 derivatives and that other heterologous DNA can be incorporated into pAg1 and pAg2 derivatives using methods well known in the art.

3. pAgIIa and pAgIIb transformation vectors

25 Vectors pAgIIa and pAgIIb were constructed by inserting the following DNA fragments into pAg1: DNA encoding β -glucuronidase, the nopaline synthase terminator fragment, the nopaline synthase (NOS) promoter fragment, a fragment of mouse satellite DNA and an *N. tabacum*

rDNA intergenic spacer sequence (IGS). The construction of pAgIIa and pAgIIb was as follows.

An *N. tabacum* rDNA intergenic spacer (IGS) sequence (SEQ. ID. NO: 98; see also GenBank Accession No. YO8422; see also Borysyuk *et al.* (2000) *Nature Biotechnology* 18:1303-1306; Borysyuk *et al.* (1997) *Plant Mol. Biol.* 35:655-660; U.S. Patent Nos. 6,100,092 and 6,355,860) was obtained by PCR amplification of tobacco genomic DNA. The IGS can be used as a targeting sequence by virtue of its homology to tobacco rDNA genes; the sequence is also an amplification promoter sequence in plants. This fragment was amplified using standard PCR conditions (*e.g.*, as described by Promega Biotech, Madison, WI, U.S.A.) from tobacco genomic DNA using the primers shown below:

NTIGS-FI

5'- GTG CTA GCC AAT GTT TAA CAA GAT G- 3' (SEQ ID No. 99) and

NTIGS-RI

5'-ATG TCT TAA AAA AAA AAA CCC AAG TGA C- 3' (SEQ ID No. 100)

Following amplification, the fragment was cloned into pGEM-T Easy to give pIGS-I. A fragment of mouse satellite DNA (Msat1 fragment; GenBank Accession No. V00846; and SEQ ID No. 101) was amplified via PCR from pSAT-1 using the following primers:

MSAT-F1

5'- AAT ACC GCG GAA GCT TGA CCT GGA ATA TCG C -3'(SEQ ID No. 102) and

MSAT-RI

5'-ATA ACC GCG GAG TCC TTC AGT GTG CA T- 3' (SEQ ID No. 103)

This amplification added a *SacII* and a *HindIII* site at the 5' end and a *SacII* site at the 3' end of the PCR fragment. This fragment was then cloned into the *SacII* site in pIGS-1 to give pMIGS-1, providing a eukaryotic

centromere-specific DNA and a convenient DNA sequence for detection via FISH.

A functional marker gene containing a NOS-promoter:GUS:NOS terminator fusion was then constructed containing the NOS promoter
5 (GenBank Accession No. U09365; SEQ ID No. 104), *E. coli* β -glucuronidase coding sequence (from the GUS gene; GenBank Accession No. S69414; and SEQ ID No. 105), and the nopaline synthase terminator sequence (GenBank Accession No. U09365; SEQ ID No. 107). The NOS promoter in pGEM-T-NOS was added to a promoterless GUS
10 gene in pBlueScript (Stratagene, La Jolla, CA, U.S.A.) using *NotI/Spel* to form pNGN-1, which has the NOS promoter in the opposite orientation relative to the GUS gene.

pMIGS-1 was digested with *NotI/Spel* to yield a fragment containing the mouse major satellite DNA and the tobacco IGS which was
15 then added to *NotI*-digested pNGN-1 to yield pNGN-2. The NOS promoter was then re-oriented to provide a functional GUS gene, yielding pNGN-3, by digestion and religation with *Spel*. Plasmid pNGN-3 was then digested with *HindIII*, and the *HindIII* fragment containing the β -glucuronidase coding sequence and the rDNA intergenic spacer, along with the Msat
20 sequence, was added to pAG-1 to form pAgIIa (SEQ ID NO: 108), using the unique *HindIII* site in pAg1 located near the right T-DNA border of pAg1, within the T-DNA region.

Another plasmid vector, referred to as pAgIIb, was also recovered, which contained the inserted *HindIII* fragment (SEQ ID NO: 108) in the
25 opposite orientation relative to that observed in pAgIIa. Thus, pAgIIa and pAgIIb differ only in the orientation of the *HindIII* fragment containing the mouse major satellite sequence, the GUS DNA sequence and the IGS sequence. The nucleotide sequence of pAgIIa is provided in SEQ. ID. NO: 109.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A eukaryotic chromosome comprising one or a plurality of *att* site(s), wherein:
 - an *att* site is heterologous to the chromosome; and
 - 5 an *att* site permits site-directed integration in the presence of lambda integrase.
2. The eukaryotic chromosome of claim 1, wherein the *att* sites are selected from the group consisting of *attP* and *attB* or *attL* and *attR*, or variants thereof.
- 10 3. The eukaryotic chromosome of claim 1 that is an artificial chromosome.
4. The eukaryotic chromosome of claim 1 that is an artificial chromosome expression system (ACes).
5. The eukaryotic chromosome of claim 4 that is predominantly
15 heterochromatin.
6. The chromosome of claim 1 that is an artificial chromosome that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
7. The chromosome of claim 1 that is a plant chromosome.
- 20 8. The chromosome of claim 1 that is an animal chromosome.
9. The chromosome of claim 7 that is a plant artificial chromosome.
10. The chromosome of claim 8 that is an animal artificial chromosome.
- 25 11. The chromosome of claim 8 that is a mammalian chromosome.
12. The chromosome of claim 11 that is a mammalian artificial chromosome.

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).

14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase
5 catalyzed recombination.

15. The *ACes* of claim 14 that contains one site.

16. The *ACes* of claim 14 that is predominantly heterochromatin.

17. The *ACes* of claim 14 that contains no more than about
30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.

10 18. The *ACes* of claim 14 that is a plant *ACes*.

19. The *ACes* of claim 14 that is an animal *ACes*.

20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.

21. The *ACes* of claim 14 that is a fish *ACes*.

15 22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli* and the R/RS
20 system of the pSR1 plasmid, or any combination thereof.

23. A method of introducing heterologous nucleic acid into a chromosome, comprising:

contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a
25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

28. The combination of claim 27, further comprising nucleic acid
15 encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase
20 system of phage Mu, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

31. The combination of claim 27, wherein a vector is the plasmid
25 pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

(a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second

5 recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

10 34. The method of claim 33, wherein the artificial chromosome is an *ACes*.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells *ex vivo*.

15 36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an *in vitro* reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

20 38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of *Int*, *IHF*, *Xis* and *Cre*, $\gamma\delta$,
25 *Tn3* resolvase, *Hin*, *Gin*, *Cin* and *Flp*.

40. The method of claim 32, wherein the recombination sites are selected from the group consisting of *att* and *lox P* sites.

41. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that removes one or more stop codons.

42. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that avoids hairpin formation.

43. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence selected from the group consisting of SEQ ID NOs:41-56:

- 10 a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:41);
- b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:42);
- c) GTTCAGCTTTCKTRTACNAACTSGB (m-attR) (SEQ ID NO:43);
- d) AGCCWGCTTTCKTRTACNAAAGTSGB (m-attL) (SEQ ID NO:44);
- e) GTTCAGCTTTYKTRTACNAAAGTSGB (m-attP1) (SEQ ID NO:45);
- 15 f) AGCCTGCTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:46);
- g) AGCCTGCTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:47);
- h) ACCCAGCTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:48);
- i) GTTCAGCTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:49);
- j) GTTCAGCTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:50);
- 20 k) GTTCAGCTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:51);
- l) AGCCTGCTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:52);
- m) AGCCTGCTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:53);
- n) ACCCAGCTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:54);
- o) GTTCAGCTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:55);
- 25 p) GTTCAGCTTCTTGTACAAAGTTGG (attP2, P3) (SEQ ID NO: 56);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G, K=G or T/U, Y=C or T/U, W=A or T/U, N=A or C or G or T/U, S=C or G, and B=C or G or T/U; and

the core region does not contain a stop codon in one or more reading frames.

44. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence
5 selected from the group consisting of a mutated att recombination site containing at least one mutation that enhances recombinational specificity, a complementary DNA sequence thereto, and an RNA sequence corresponding thereto.

45. The method of claim 33, wherein the vector comprising the
10 second site further encodes at least one selectable marker.

46. The method of claim 45, wherein the marker is a promoterless marker, which, upon recombination is under the control of a promoter and is thereby expressed.

47. The method of claim 46, wherein the first recombination site
15 is attP and is in the sense orientation prior to recombination.

48. The method of claim 46, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic, fluorescent, or capable of being bound by an antibody and FACs sorted.

20 49. The method of claim 48, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase (hisD).

50. A cell comprising, the chromosome of claim 1.

25 51. The cell of claim 50, wherein the cell is a nuclear donor cell.

52. The cell of claim 50, wherein the cell is a stem cell.

53. The stem cell of claim 52, wherein said stem cell is human and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

54. The cell of claim 50, wherein the cell is mammalian.
55. The cell of claim 54, wherein the mammal is selected from the group consisting of humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, and horses.
- 5 56. The cell of claim 50, wherein the cell is a plant cell.
57. A cell comprising the platform ACes of claim 14.
58. The cell of claim 57, wherein the cell is a nuclear donor cell.
59. The cell of claim 57, wherein the cell is a stem cell.
60. The stem cell of claim 59, wherein said stem cell is human
- 10 and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.
61. A human mesenchymal cell comprising an artificial chromosome.
62. The human mesenchymal cell of claim 61, wherein said
- 15 artificial chromosome is an ACes.
63. The human mesenchymal cell of claim 62, wherein the ACes is a platform-ACes.
64. A method for introducing heterologous nucleic acid into the mesenchymal cell of claim 63, comprising:
- 20 (a) introducing into the cell of claim 63, wherein the platform-ACes has a first recombination site, a vector comprising at least a second recombination site and the heterologous nucleic acid;
- (b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between
- 25 the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the platform-ACes within the mesenchymal cell.
65. A lambda-intR mutein comprising a glutamic acid to arginine change at position 174 of wild-type lambda-intR.

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence
10 of a IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an
25 embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

76. The method of claim 72, wherein the transgenic animal is a fish, insect, reptile, amphibians, arachnid or mammal.

77. The method of claim 72, wherein the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection,
5 microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

78. A transgenic animal produced by the method of claim 72.

79. A cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated
10 throughout the endogenous chromosomes.

80. A method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into the cell line of claim 79, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of
15 the heterologous recombination sites within the cell's chromosomal DNA; and isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within.

81. A library of cells useful for genomic screening, said library
20 comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein.

82. The library of cells of claim 81, wherein the cells of the library are from a different species than the chromosomal nucleic acid
25 within the *ACes*.

83. A method of making one or more cell lines, comprising
a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,

b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;

5 c) isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;

d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;

10 e) selecting different cells having mutually exclusive *ACes* therein and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

84. The method of claim 23, wherein the nucleic acid molecule with a recombination site is a PCR product.

15 85. Method of claim 23 wherein the recombinase is a protein and the recombination event occurs in vitro.

86. The method of claim 33, wherein the vector is a PCR product comprising a second recombination site.

20 87. The lambda-intR mutein of claim 65, wherein the mutein further comprises an amino acid signal for nuclear localization.

88. The lambda-intR mutein of claim 65, wherein the mutein further comprises an epitope tag for protein purification.

89. A modified iron-induced promoter comprising SEQ ID NO:128.

25 90. A plasmid or expression cassette comprising the promoter of claim 89.

91. A vector, comprising:
a recognition site for recombination; and

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:

- 5 introducing a vector of claim 91 into a cell;
 growing the cells; and
 selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein sufficient portion of the
10 vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:

- 15 contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

 the reporter *ACes* comprises one or a plurality of reporter constructs;

- a reporter construct comprises a reporter gene in operative linkage
20 with a regulatory region responsive to test or known compounds; and
 detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively
25 linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent
5 proteins.

113. The method of claim 109, wherein the *ACes* comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cell.

122. The method of claim 109, wherein the cell is in an animal;
and the method comprises whole-body imaging to monitor expression of
5 the reporter in the animal.

123. A reporter ACes comprises one or a plurality of reporter
constructs, wherein the reporter construct comprises a reporter gene in
operative linkage with a regulatory region responsive to test or known
compounds.

Generation of ACes for Platform Chromosome Engineering

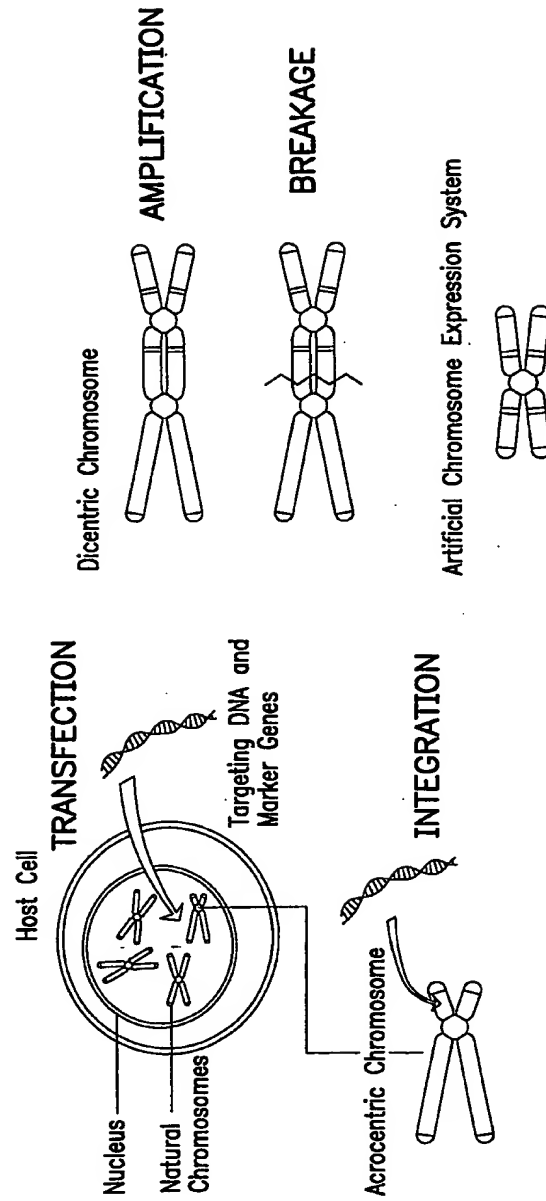


FIG. 1

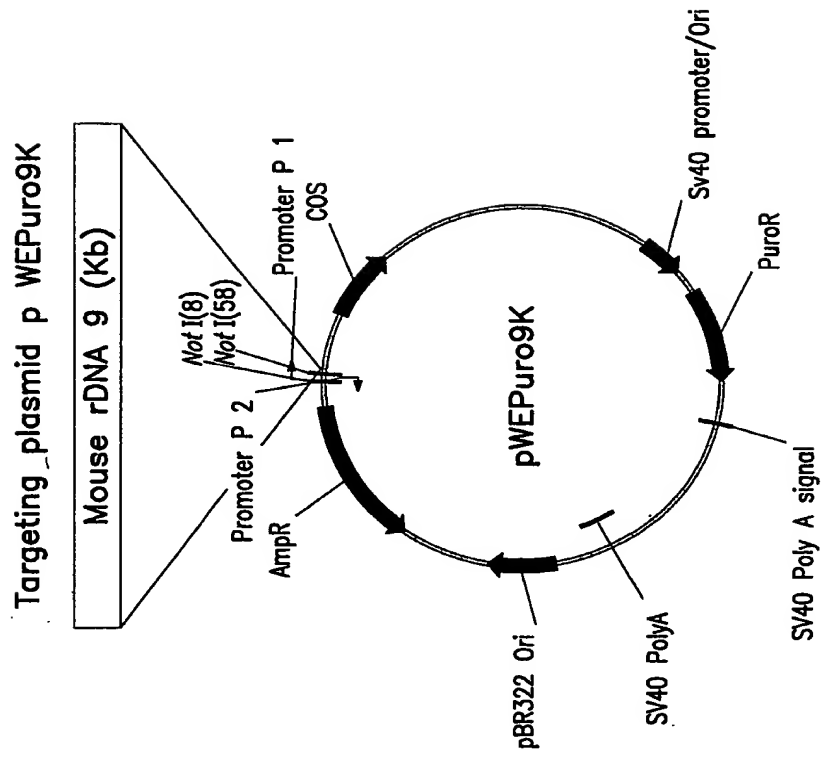


FIG. 2

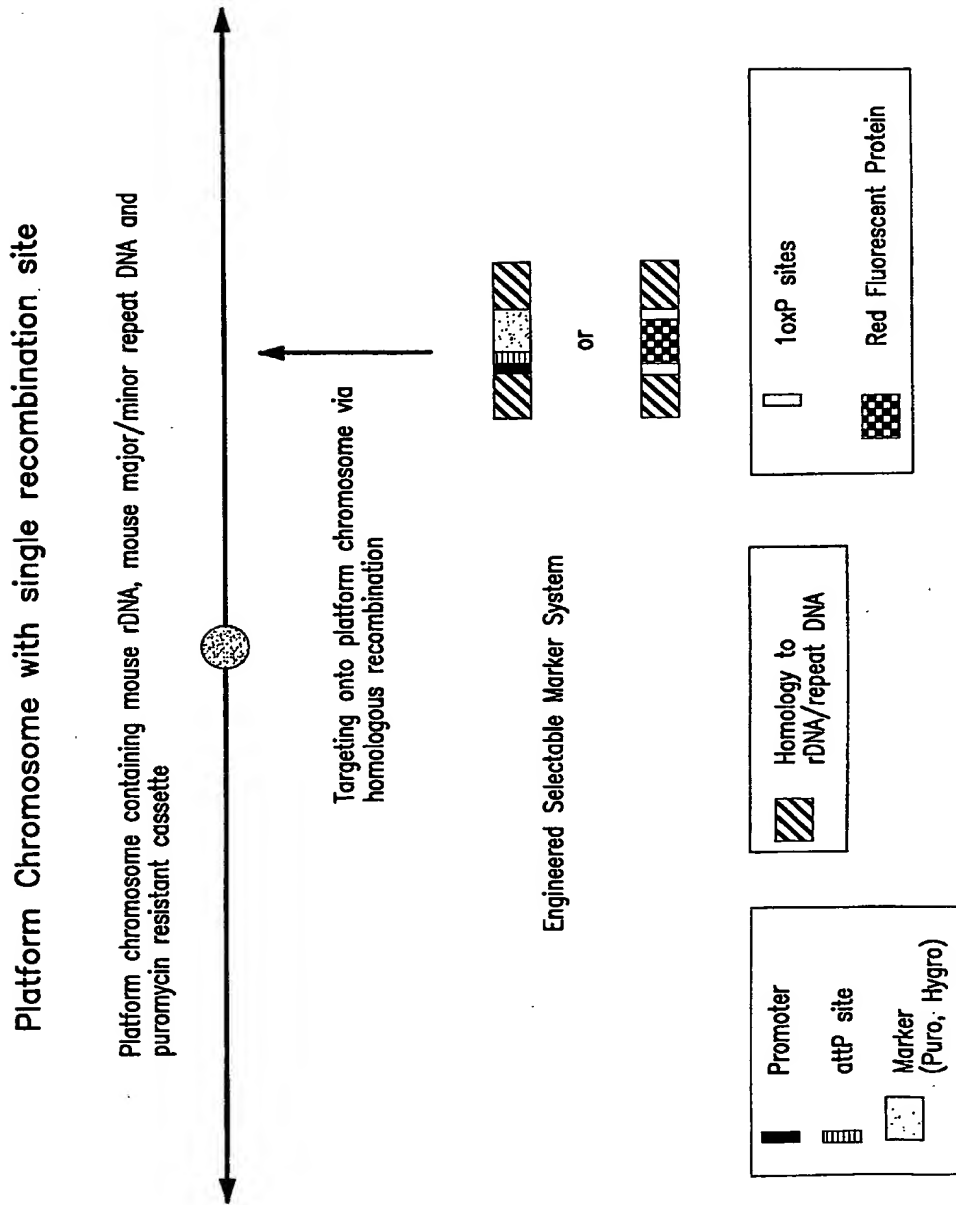


FIG. 3

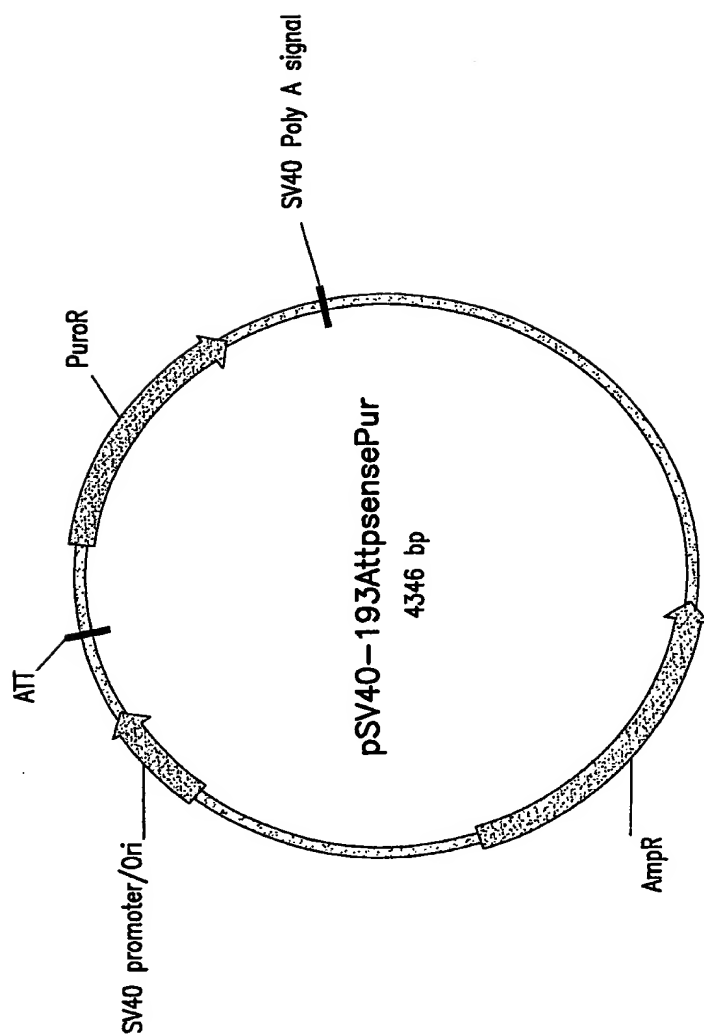


FIG. 4

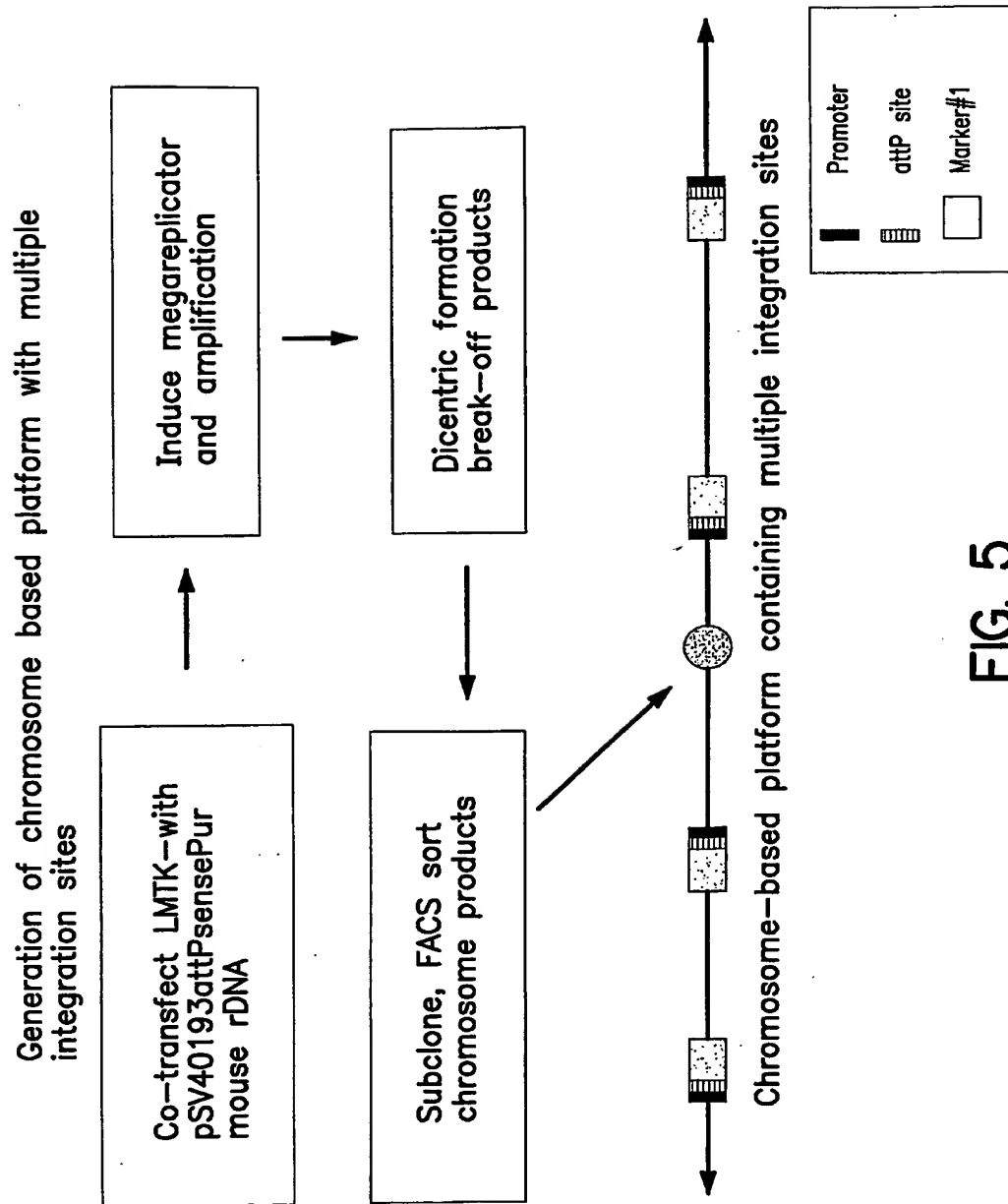


FIG. 5

λ integrase recombination



	Core Region
attP	CAGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attB	CTGCTTTT <u>T</u> TATACTAA <u>C</u> TTG
attL	CTGCTTTT <u>T</u> TATACTAA <u>G</u> TTG
attR	CAGCTTTT <u>T</u> TATACTAA <u>C</u> TTG

FIG. 6

λ INT recombination on artificial chromosome

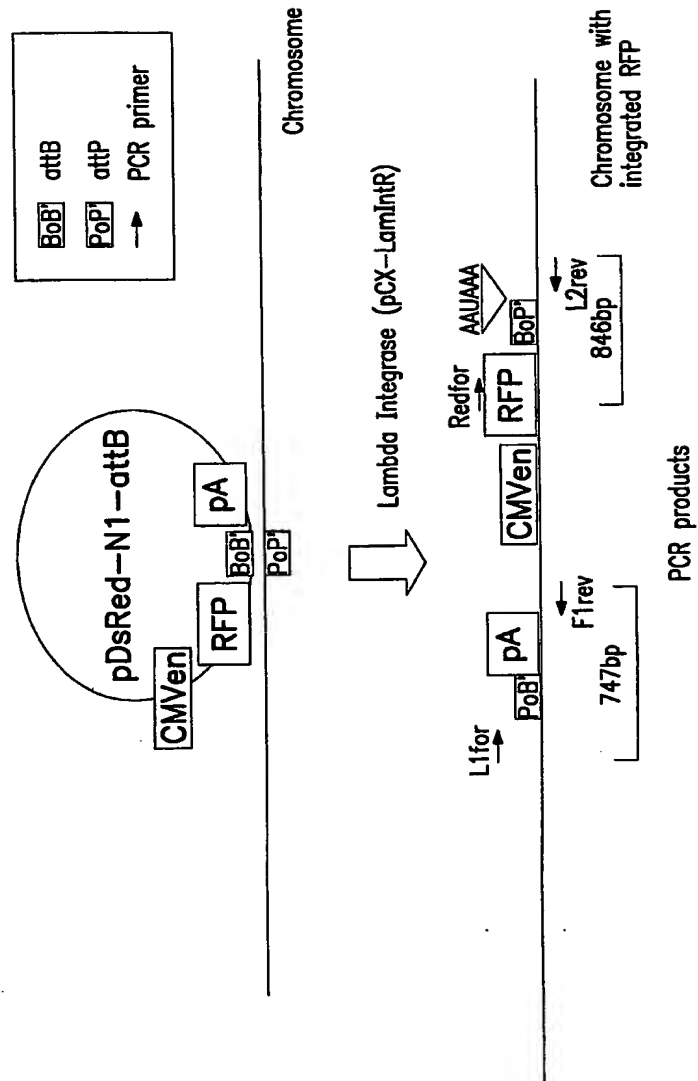


FIG. 7

pCXLamIntR integrase expression vector

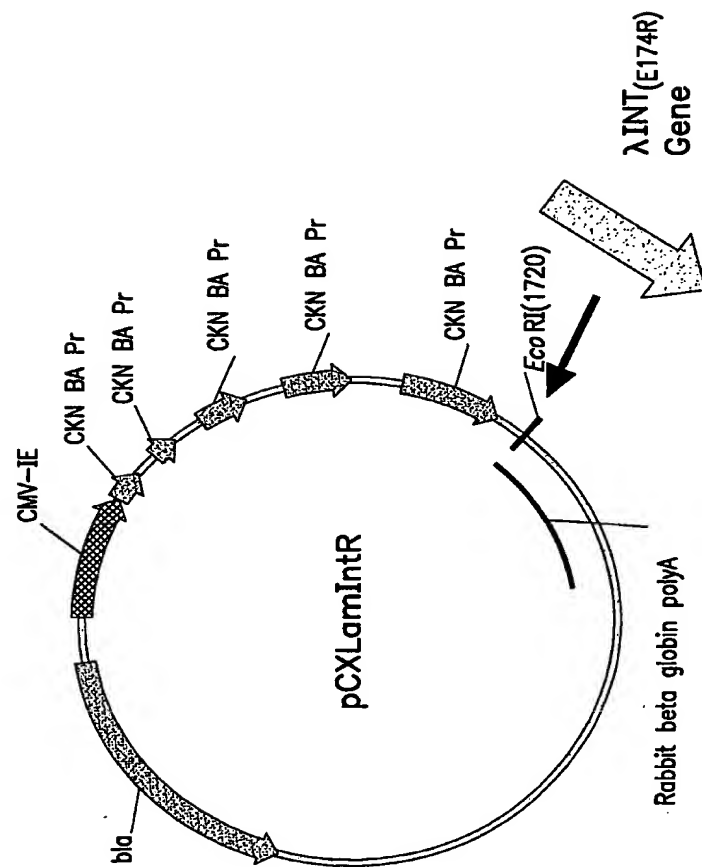


FIG. 8



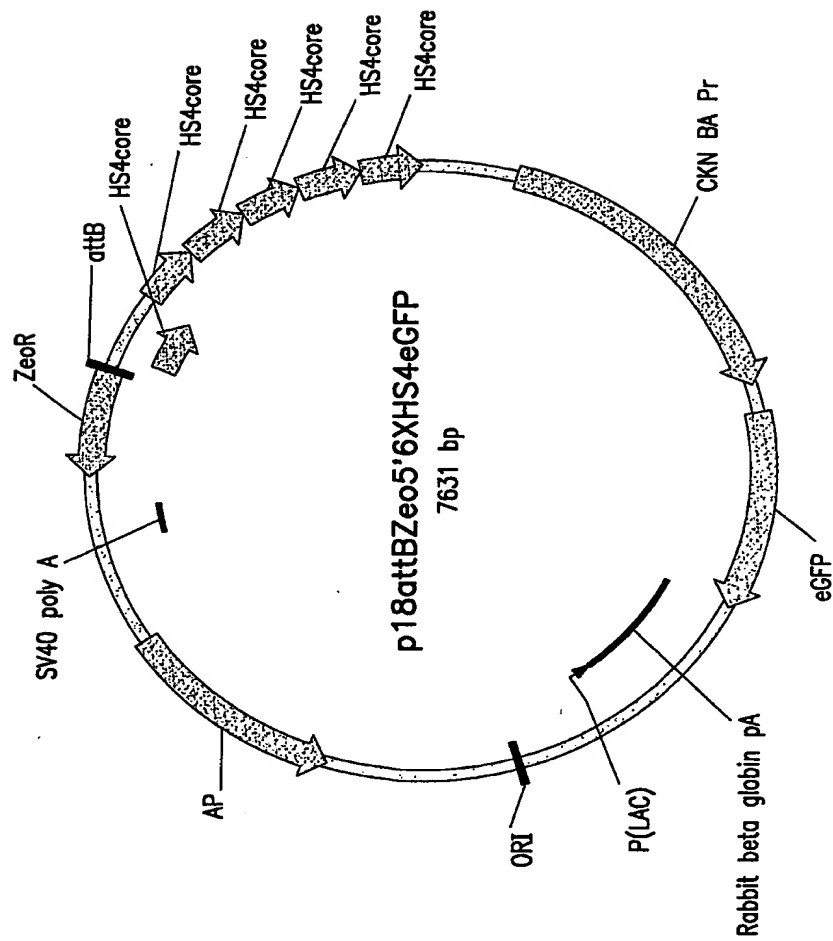


FIG. 10

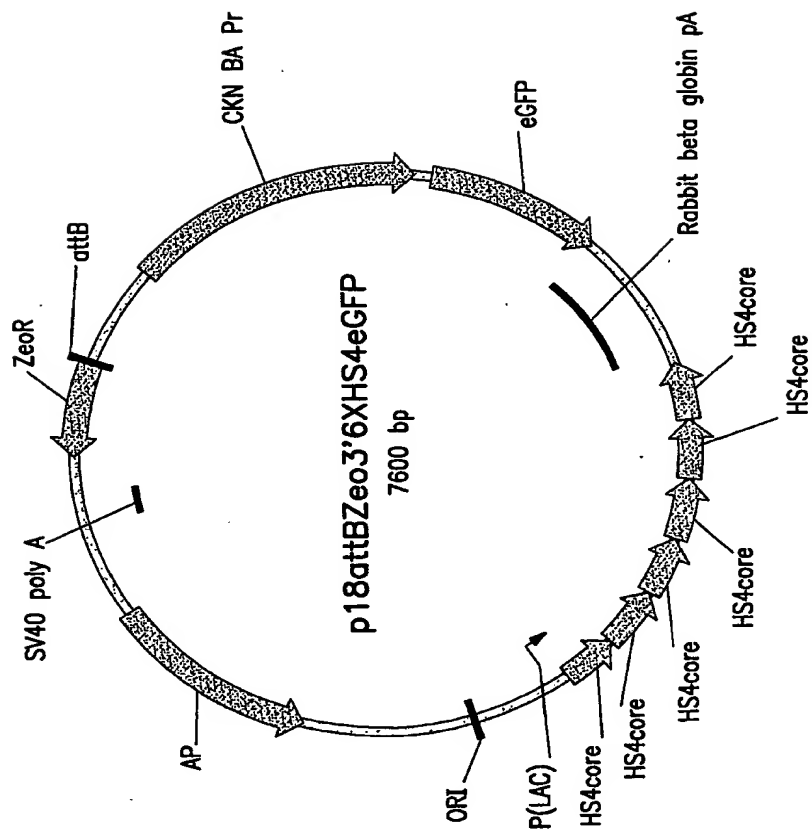


FIG. 11

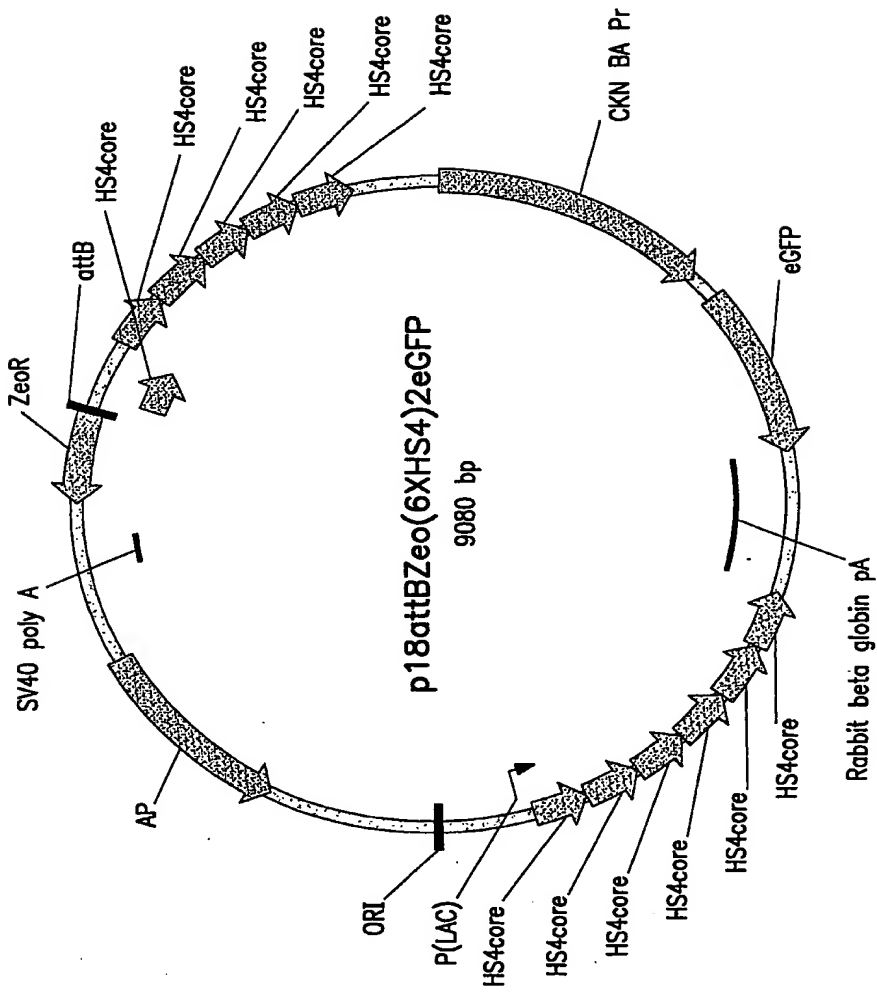


FIG. 12

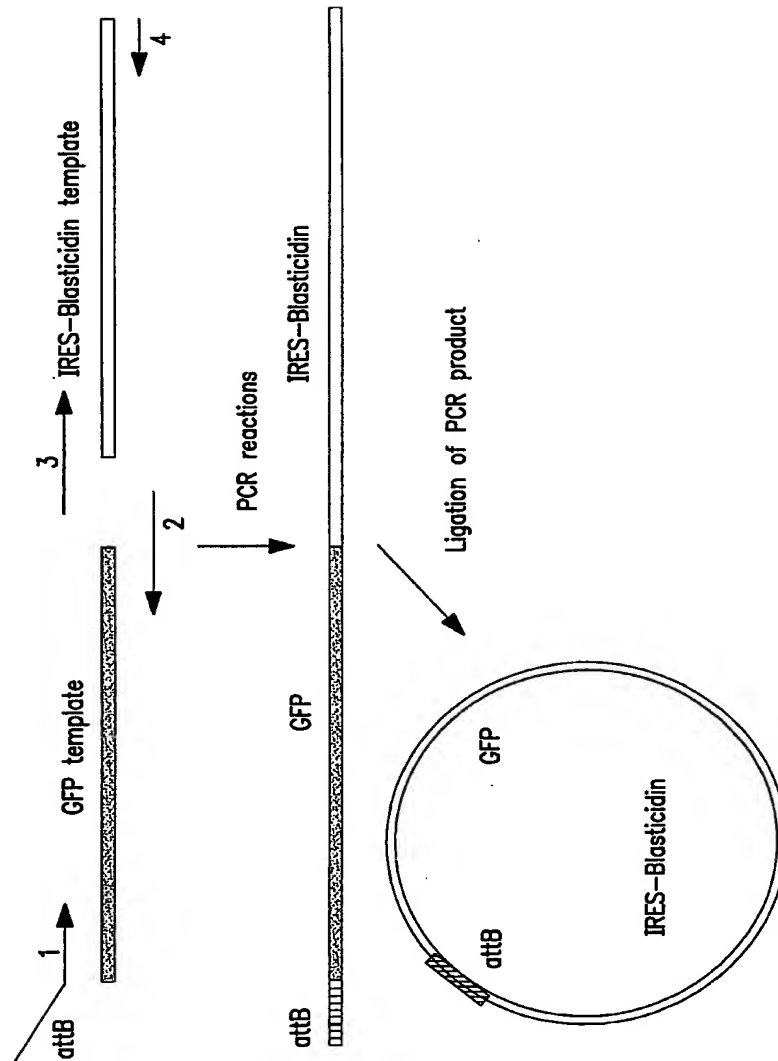


FIG. 13

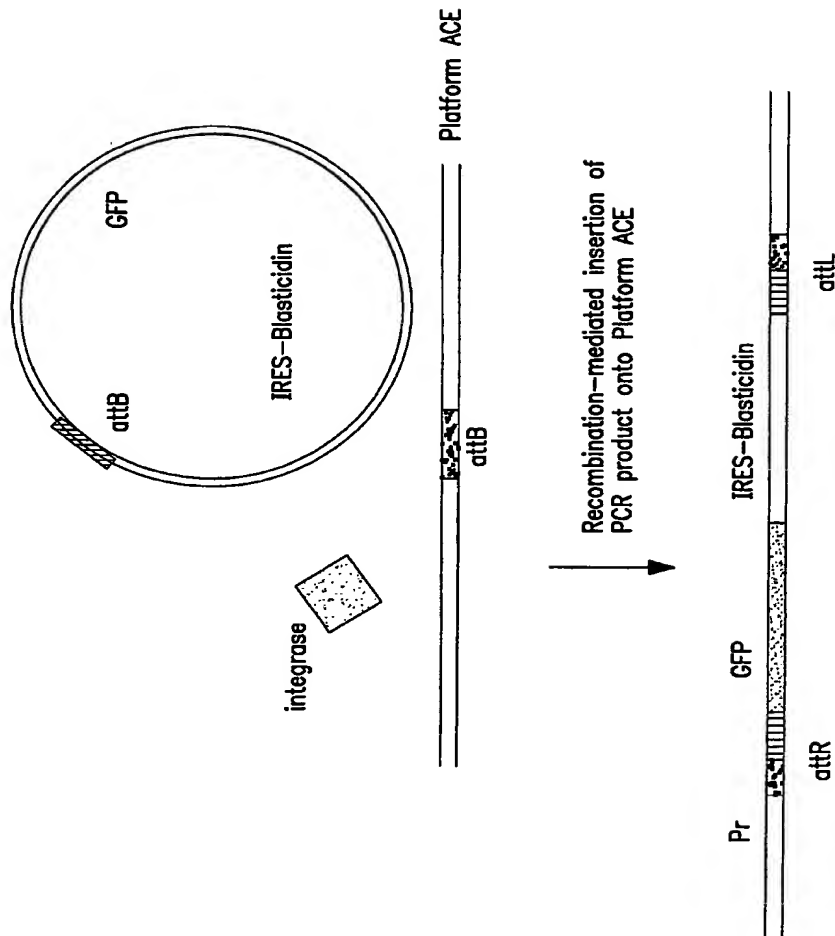


FIG. 14

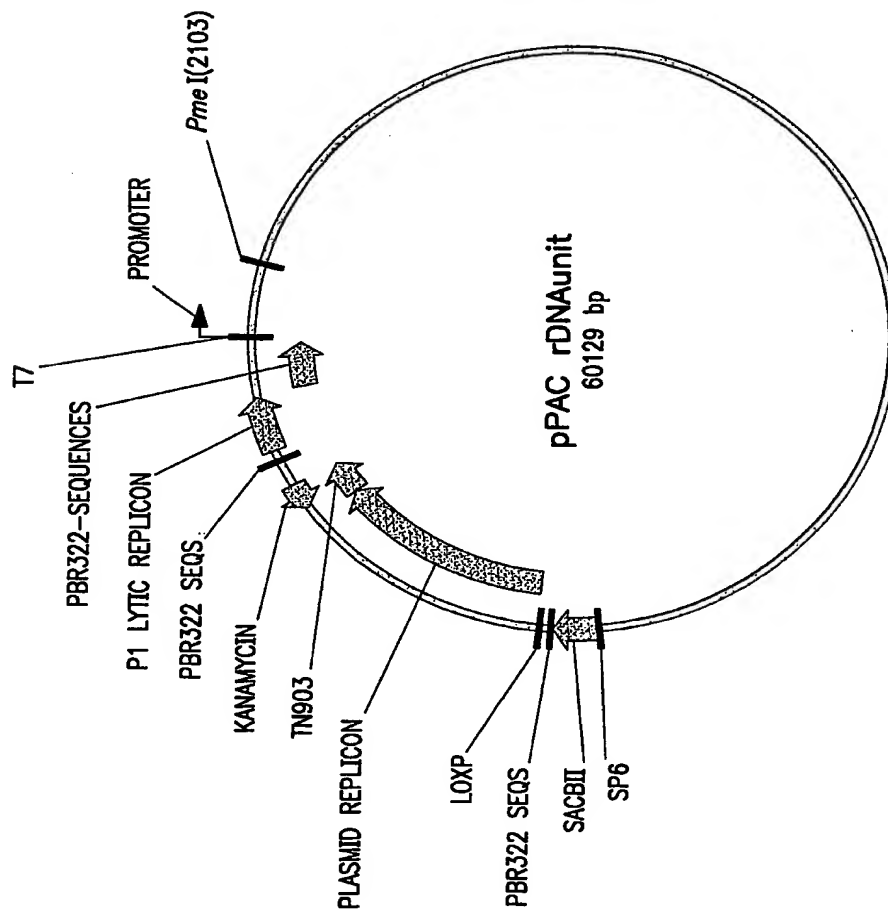


FIG. 15

SEQUENCE LISTING

<110> CHROMOS MOLECULAR SYSTEMS, INC.

Perkins, Edward
Perez, Carl
Lindenbaum, Michael
Greene, Amy
Leung, Josephine
Fleming, Elena
Stewart, Sandra
Shellard, Joan

<120> CHROMOSOME-BASED PLATFORMS

<130> 24601-420PC

<140> Not Yet Assigned

<141> Herewith

<150> 60/294,758

<151> 2001-05-30

<150> 60/366,891

<151> 2002-03-21

<160> 129

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: attPUP

<400> 1

ccttgcgcta atgctctgtt acagg

25

<210> 2

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: attPDWN

<400> 2

cagaggcagg gagtgggaca aaattg

26

<210> 3

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: Lamint 1

<400> 3

ttcgaattca tgggaagaag gcgaagtcac gagcg

35

<210> 4

<211> 34

<212> DNA

<213> Artificial Sequence

<220>
 <223> Primer: Lamint 2

<400> 4
 ttccaattct tatttgattt caattttgtc ccac 34

<210> 5
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer

<400> 5
 cggacaatgc ggttgtgcgt 20

<210> 6
 <211> 46
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 6
 cgcgagcaa aatctagagt aaggagatca agacttacgg ctgacg 46

<210> 7
 <211> 46
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> LambdaINTER174rev

<400> 7
 cgtcagccgt aagtcttgat ctcttactc tagattttgc tgcgcg 46

<210> 8
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attB1

<400> 8
 tgaagcctgc ttttttatac taacttgagc gaa 33

<210> 9
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attB2

<400> 9
 ttcgctcaag ttagtataaa aaagcaggct tca 33

<210> 10
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Primer: attPdwn2

<400> 10
tcttctcggg cataagtcgg acacc 25

<210> 11
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVen

<400> 11
ctcacgggga tttccaagtc tccac 25

<210> 12
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:attPdwn

<400> 12
cagaggcagg gagtgggaca aaattg 26

<210> 13
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:CMVEN2

<400> 13
caactccgcc ccattgacgc aaatg 25

<210> 14
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L1

<400> 14
agtatcgccg aacgattagc tcttca 26

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:F1 rev

<400> 15
gccgatttcg gcctattggt taaa 24

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:RED

<400> 16
ccgccgacat ccccgactac aagaa

25

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer:L2rev

<400> 17
ttccttcgaa ggggatccgc ctacc

25

<210> 18
<211> 22118
<212> DNA
<213> Mus musculus

<300>
<308> GenBank X82564
<309> 1996-04-09

<400> 18	gaattccccct	atccctaatac	cagattgggtg	gaataacttg	gtatagatgt	ttgtgcatta	60
	aaaaccctgt	aggatcttca	ctctagggtca	ctgttcagca	ctggaacctg	aattgtggcc	120
	ctgagtata	gggtcctggga	catatgcagt	tctgcacaga	cagacagaca	gacagacaga	180
	cagacagaca	gacagacgtt	acaaacaaac	acgttgagcc	gtgtgccaac	acacacacaa	240
	acaccactct	ggccataatt	attgaggacg	ttgatttatt	attctgtgtt	tgtgagtctg	300
	tctgtctgtc	tgtctgtctg	tctgtctgtc	tatcaaacca	aaagaaacca	aacaattatg	360
	cctgcctgcc	tgccctgcctg	cctacacaga	gaaatgattt	cttcaatcaa	tctaaaacga	420
	cctcctaagt	ttgccttttt	tctctttctt	tatcttttct	ttttttcttt	tcttcttctt	480
	tccttccttc	cttccttcctt	tccttccttt	ctttctttct	ttctttcttt	cttactttct	540
	ttctttctct	cttacatttta	ttcttttcat	acatagtttc	ttagtgttaag	catccctgac	600
	tgtcttgaag	acacttttga	ggcctcaatc	ctgttaagagc	cttcctctgc	ttttcaaatg	660
	ctggcatgaa	tgttgtacct	cactatgacc	agcttagtct	tcaagtctga	gttactggaa	720
	aggagttcca	agaagactgg	ttatattttt	cattttattat	tgcattttaa	ttaaaaattta	780
	atttcaccaa	aagaattttag	actgaccaat	tcagagctctg	ccgttttaaaa	gcataaggaa	840
	aaagtaggag	aaaadacgtga	ggctgtctgt	ggatggctga	ggctgcttta	gggagcctcg	900
	tcaccattct	gcacttgcaa	accggggccac	tagaaccggg	tgaaggagga	aaccaaagcg	960
	acctggaaac	aatagggtcac	atgaaggcca	gccacctcca	tcttggtgtg	cgggagttca	1020
	gttagcagac	aagatggctg	ccatgcacat	gttgctcttc	agcttggtga	gggtcaagta	1080
	caaccgagtc	acagaacaag	gaagtataca	cagtgaagttc	caggtcagcc	agagtttaca	1140
	cagagaaacc	acatcttgaa	aaaaacaaaa	aaataaatta	aataaatata	atttaaaaat	1200
	ttaaaaatag	ccgggagtgga	tgccgcatgt	ctttaatccc	agctctcttc	aggcagagat	1260
	gggaggattt	ctgagtttga	ggccagcctg	gtctgcaaag	tgagttccag	gacagtcagg	1320
	gctatacaga	gaaaccctgt	cttgaaaact	aaactaaatt	aaactaaaact	aaactaaaaa	1380
	aatataaaat	aaaaatttta	aagaatttta	aaaaactaca	gaaatcaaac	ataagcccac	1440
	gagatggcaa	gtaactgcaa	tcatagcaga	aatattatac	acacacacac	acacagactc	1500
	tgtcataaaa	tcctaatgtgc	cttcatgatg	atcaaatttc	gatagtcagt	aatactagaa	1560
	gaatcatatg	tctgaaaata	aaagccagaa	ccttttctgc	ttttgttttc	ttttgccccca	1620
	agatagggtt	tctctcagtg	tatccctggc	atccctgcct	gggaacttct	ttgtagggtt	1680
	ggtagcctca	aaactcagaga	ggtcctctct	gcctgcctgc	ctgctgcctc	gcctgcctgc	1740
	ctgcctgcct	gcctgcctca	cttcttctgc	caccacacaca	accgagtcga	accatggatc	1800
	tttatttctt	tctctttctc	tcttctttct	ttctttcttt	ctttctttct	ttctttcttt	1860
	ctttctttct	ttcttatttc	attagttttc	aatgtaagt	tgtgtttgtg	ctctatctgc	1920
	tgccatagg	cctgcttgcc	aggagagggc	aacagaacct	aggagaaacc	accatgcagc	1980
	tcctgagaat	aagtgaaaaa	acaacaaaaa	aaggaaattc	taatcacata	gaatgtagat	2040
	atatgccgag	gctgtcagag	tgctttttta	ggcttagtgt	aagtaatgaa	aattgttgtg	2100
	tgtcttttat	ccaaacacag	aagagaggtg	gctcggcctg	catgtctgtt	gtctgcaggt	2160
	agaccaggct	ggccttgaaac	acattaatct	gtctgcctct	gcttccctaa	tgctgcgatt	2220
	aaaggcatgt	gccaccactg	cccggaactga	tttcttcttt	tttttttttt	tggaataaac	2280
	ctttctttct	ttttctctct	ctctttcttc	cttcttctct	ttctttctat	tctttttttc	2340
	tttctttttt	tttttttttt	ttttttttta	aatgtgccta	agggttaaagg	tgtgctccac	2400
	aattgcctca	gctctgctct	aattctcttt	aaaaaaaaac	aaacaaaaaa	aaaacaaaaa	2460
	cagtatgtat	gtatgtatat	ttagaagaaa	tactaatcca	ttataaactc	tttttctcta	2520
	aaattcatgt	cattcttgtt	ccacaaagtg	agttccagga	cttaccagag	aaaccctgtg	2580

ttcaaatttc	tgtgtttcaag	gtcaccctgg	cttacaaagt	gagttccaag	tccgataggg	2640
ctacacagaa	aaaccatata	tcagaaaaaa	aaaaagttcc	aaacacacac	acacacacac	2700
acacacacac	acacacacac	acacacacac	acacacacag	cgcgcgcgcg	cgatgagggg	2760
aagtcgtgcc	taaaaataaat	atTTTTtctgg	ccaaagttaa	agcaaatcac	tatgaagagg	2820
tactcctaga	aaaaataaat	acaaacgggc	TTTTtaataca	ttccagcact	gttttaattt	2880
aactctgaat	ttagtcttgg	aaaagggggc	gggtgtgggt	gagtgagggc	gagcgagcag	2940
acgggcgggc	gggcgggtga	gtggccggcg	gcgtgtggcag	cgagcaccag	aaaacaacaa	3000
accccaagcg	gtagagtgtt	ttaaaaatga	gacctaaatg	tggtggaacg	gaggtcgccg	3060
ccacctcct	cttccactgc	ttagatgctc	ccttccccct	actgtgctcc	cttccccctaa	3120
ctgtgcctaa	ctgtgcctgt	tccctcacc	cgctgattcg	ccagcgacgt	actttgactt	3180
caagaacgat	tttgcctgtt	ttcacccgct	cctgtcatat	tttcgttttt	gggtgcccca	3240
gtctagcccc	ttcgctatgt	tcgggcggga	cgatggggac	cgtttgtgcc	actcggggaga	3300
agtgggtggg	gggtacgtg	ctccgtcggt	cggtgtggag	tgccgggaacc	tgagctcggg	3360
agaccctccg	gagagacaga	atgagtgaat	gaatgtggcg	gcgcgtgacg	gatctgtatt	3420
ggtttgtatg	gttgatcgag	accattgtcg	ggcgacacct	agtggtgaca	agtttcggga	3480
acgctccagg	cctctcaggt	tggtgacaca	ggagagggaa	gtgcctgtgg	tgaggcgacc	3540
agggtgacag	gaggccgggc	aagcaggcgg	gagcgtctcg	gagatgggtg	cgtgtttaag	3600
gacggtctct	aacaaggagg	tcgtacaggg	agatggccaa	agcagaccga	gttgctgtac	3660
gcccttttgg	gaaaaatgct	agggttgggt	gcaacgttac	taggtcgacc	agaaggctta	3720
agtcctaccc	ccccccccct	TTTTTTTTTT	tttctccag	aagccctctc	ttgtccccgt	3780
caccgggggc	accgtacatc	tgaggccgag	aggacgcgat	gggcccggct	tccaagccgg	3840
tgtggctcgg	ccagctggcg	cttcgggtct	TTTTTTTTTT	TTTTTTTTTT	tttctctcca	3900
gaagccttgt	ctgtcgctgt	caccgggggc	gctgtacttc	tgaggccgag	aggacgcgat	3960
gggccccggc	ttccaagccg	gtgtggctcg	gccagctgga	gcttcgggtc	TTTTTTTTTT	4020
TTTTTTTTTT	TTTTTTTTTc	cagaagccct	gtctgtcgct	gtcaccgggg	gcgctgtact	4080
tctgaggccg	agaggacgcg	atgggtcgcc	ttccaagccg	atgtggcggg	gccagctgga	4140
gcttcgggtt	TTTTTTTTTc	ctccagaagc	cctctcttgt	ccccgtcacc	gggggcgctg	4200
tacttctgag	gccgagagga	cgtgatgggc	ccgggttcca	ggcggatgtc	gccccgtcag	4260
ctggagcctt	ggatcttttt	TTTTTTTTTT	cctccagaag	ccctctcttg	ttcccgtcac	4320
cgggggcacc	ttacatctga	gggcgagagg	acgtgatggg	tcgggcttcc	aagccgatgt	4380
ggcggggcca	gctggagcct	cggttttttt	TTTTTctctc	cagaagccct	ctcttgtccc	4440
cgtcaccggg	ggcgctgtac	ttctgaggcc	gagaggacgt	gatgggcccc	ggttccaggc	4500
ggatgtcgcc	cggtcagctg	gacgtttgga	tcattttttt	TTTTTccctc	agaagccctc	4560
tcttgtcccc	gtcacccggg	gcaccgtaca	tctgaggccg	agaggacacg	atgggcctgt	4620
cttccaagcc	gatgtggccc	ggccagctgg	agcttcgggt	cttttttttt	TTTTTctctc	4680
cagaagccct	gtctgtcgct	gtcaccgggg	gcgctgtact	tctgaggccg	agaggacgcg	4740
atgggcccgg	cttccaagcc	gggtgtggctc	ggccagctgg	agcttcgggt	cttttttttt	4800
TTTTTTTTTT	ttcctccaga	aaccttgtct	gtcgctgtca	cccggggcgc	ttgtacttct	4860
gatgccgaga	ggacgcgatg	ggcccgtctt	ccagggccgat	gtggcccggg	cagctggagc	4920
tttggatctt	TTTTTTTTTT	TTTTTctcca	gaagccctct	cttgtccccg	tcaccggggg	4980
caccttacat	ctgaggccta	gaggacacga	tgggcccggg	ttccaggccg	atgtggcccc	5040
gtcagctgga	gctttggatc	TTTTTTTTTT	TTTTTctcca	gaagccctct	tgtccccctc	5100
accggtggga	ctgtacatct	gaggcgagga	ggacattatg	ggcccggcct	ccaatccgat	5160
gtggcccggc	cagctggagc	tttggatctt	attttttttt	taattttttc	ttccagaagc	5220
cctcttgtcc	ctgtcaccgg	tgccacggta	catctgaggc	cgagaggaca	ttatgggccc	5280
ggcttccagg	ccgatgtggc	ccggctcagct	ggagctttgg	atcttttttt	TTTTTTTTTc	5340
TTTTTctctc	agaagccctc	tctgtccctg	tcaccggggg	ccctgtacgt	ctgaggccga	5400
gggaaagcta	tgggcgcggt	tttctttcat	tgacctgtcg	gtcttatcag	ttctccgggt	5460
tgtcagggtc	gaccagtgtg	tccttttgagg	tcgggttctt	ttcgttatgg	gggtcattttt	5520
gggcccacct	cccaggatat	acttccaggc	gtcgttgctc	gcctgtcact	ttcctccctg	5580
tctcttttat	gcttgtgata	TTTTctatct	gttccattatg	gacctggaga	taggtactga	5640
cacgctgtcc	tttccctatt	aacactaaag	gacactataa	agagaccctt	tcgatttaag	5700
gctgttttgc	ttgtccagcc	tattcttttt	actggcttgg	gtctgtcgcg	gtgacctgaag	5760
ctgtccccga	gccacgcttc	ctgctttccc	gggagctgtg	cttgctgtgtg	cttgctgtgg	5820
gcagcttgtg	acaactgggc	gctgtgactt	tgctgcgtgt	cagacgtttt	tcccgatattc	5880
cccagggtgt	cgttgtcaca	cctgtcccgg	ttgggaatgg	ggagccagct	gtgggttgagg	5940
gccaccttat	ttcggctcac	TTTTTTTTTT	TTTTTctctc	ttggagtccc	gaacctccgc	6000
tcttttctct	tcccggctct	tcttccacat	gcctcccagag	tgcatttctt	tttgtttttt	6060
ttcttttttt	TTTTTTTTTT	ttggggagggt	ggagagtccc	gagtacttca	ctcctgtctg	6120
tggtgtccaa	gtgttcacgc	caagtgccct	ccagagtgcac	TTTTTTTTTg	ggcagtcgct	6180
cgttgtgttc	tcttgtcttg	tgtctgcccg	tatcagtaac	tgtcttgccc	cgcgtgtaag	6240
acattcctat	ctcgttgtgt	tctcccgatt	gcgcgtcgtt	gctcactctt	agatcgatgt	6300
ggtgctccgg	agttctcttc	gggcccaggc	caagcccgcc	caggcgaggg	acggacattc	6360
atggcgaaatg	gcgcccgctc	ttctcgttct	gccaagcggg	cctcgtctct	ccaccctatc	6420
cgtctgcccg	tggtgtgtgg	aaggcagggg	tgccgctctc	cggcccgcag	ctgcccccg	6480
cgcacttttc	tcagtgggtc	gcgtgggtct	tgtggatgtg	tgaggcgccc	ggttgtgccc	6540
tcacgtgttt	cactttgggt	gtgtctcgct	tgaccatgtt	cccagagtcg	gtggatgtgg	6600

cgggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgtgtttct	tgtaagcgtc	6660
gaggtgctcc	tggagcgttc	caggtttgtc	tcctagggtc	ctgottctga	gctggtggtg	6720
gcgctcccca	ttccctgggtg	tgccctccggt	gctccgtctg	gctgtgtgccc	ttcccggttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtagggtc	ggtgctgagc	6900
ggtcgcggct	gggggttgaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccggtt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgccggtg	cgcccgcgtg	7080
ttggtcttct	ggtttccctg	tgtgtctcgtc	gcatgcatcc	tcctcgggtg	gcccgggctc	7140
gtcgggggtt	tgggtccgtc	ccgcccctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaagggtg	cgggcttctt	acggtctcga	gggggtctctc	ccgaatgggtc	ccctggaggg	7260
ctcgccccct	gaccgcctcc	cgcgcgcgca	cggtttgtctc	tcctgtctac	cgccggcccgc	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtggttggt	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttcctcc	cctcctgagg	gcgcgcgtgc	ggacggggtg	7500
tgggtaggcg	acgggtgggt	cccgggtccc	caccgctctt	cccgtgcctc	acccgtgcct	7560
tcgctcgcgt	gcgtccctct	cgctcgcgtc	cacgactttg	gccgctcccg	cgacggcgcc	7620
ctgcgcgcgc	cgtggtgcgt	gctgtgtgct	tcctgggctg	tgtggttggt	tcgctccgcc	7680
cccccttccc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcgggagtc	tccttcccc	7740
cctcggggtc	gagaggggtc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacgggacc	7800
gttctgtggg	agaacggctg	ttggcccgct	ccggcgcgac	gtcggacgtg	gggacccact	7860
gccgctcggg	ggtcttcgtc	ggtaggcatc	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cgggtgcgcc	tcctcgggct	cccggggggc	cgctggtttt	cgggtcgggt	cggcgctgca	7980
ggtgtggtgg	gactgctcag	gggagtgggt	cagtgtgatt	cccgcgggtt	ttgcctcgcg	8040
tgccctgacc	ggtccgacgc	ccgagcggtc	tcctgggtccc	ttgtgaggac	ccccctccgg	8100
gagggggccc	tttcggccgc	ccttgccgtc	gtcgcgggcc	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgcgcga	gccggtcttt	tttctctct	ccccctctct	cctctgactg	8220
accggtggcc	gtgctgtcgg	accccccgca	tggggggcgc	cgggcacgta	cgcgctccggg	8280
cgggtcacccg	ggtcttgggg	ggggggccgag	gggtaagaaa	gtcggctcgg	cgggcgggag	8340
gagctgtggt	ttggagggcg	tcccggcccc	gcggccgtgg	cgggtgtctg	cgcggtcttg	8400
gagagggctg	cgtgcgaggg	gaaaaggttg	ccccgcgagg	gcaaagggaa	agaggctagc	8460
agtggctcatt	gtcccgcagc	tgtggtgggt	tgttgggcga	ggtgcgtctg	gggggctcgt	8520
cgggcccctg	cgtccgtcgg	gaaggcgcgt	ggttgggccct	gccggagtgc	cgagggtggg	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgggtg	tggcgggtggg	ggctccgggtc	8640
gatgtctacc	tcctctctcc	cgaggtctca	ggccttctcc	gcgcgggctc	tcggccctcc	8700
cctcgttctc	ccctctcgcg	gggttcaagt	cgctcgtcga	cctccccctc	tcgctccttc	8760
catctctcgc	gcaatggcgc	cgcccgagtt	cacgggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcccgggg	ctggccgctg	tcgggtctct	cctgcccgac	ccccgttggc	gtggtcttct	8880
ctgcgcggct	tcgcggactc	ctggcttctc	ccggagggtc	agggggcttc	ccgggtcccc	8940
gaogttgcgc	ctcgtctcgt	tgtgcttggt	ggggggccgc	tgccggctcc	gcccgcctcg	9000
gagccctcgc	cgcacccgcc	ggtgtgcggt	ttcgcgcgcg	ggtcagttgg	gccctggcgt	9060
tgtgtcgcgt	cgggagcgtg	tcgcctcgc	ggcggttaga	cgccgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgcccccg	ccgacccccg	cctgcccgtc	ccgggtgggtg	9180
tcgttggtgt	ggggagtgaa	tggtgctacc	ggatattccc	tcgccggtgg	tttgactgtc	9240
tcgcccgggt	cgcgcttctc	tttcggccaa	ccccacgcgc	aaccaccac	cctgctctcc	9300
cggcccgggtg	cggctcgacgt	tcgggtctct	ccgatgccga	gggggttcggg	atttgtgccc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcgggagagc	tgtcccgggg	cgacgctcgg	9420
ggttggtttg	ccgcgtgcgt	gtgctcgcgg	acgggttttg	tcggaccccg	acggggtcgg	9480
tcgggcgcga	tgcaactctc	cgttccgcgc	gagcgcgcgc	ccgggtcacc	cccgggttgt	9540
cctcccgcga	ggctctccgc	cgcgcgcgcc	tcctctcctc	ctctcgcgct	ctctgtccc	9600
cctgggtcctg	tcacaccccc	gacgctccgc	tcgcgcttcc	ttacctgggt	gatcctgcca	9660
ggtagcataat	gcttgtctca	aagattaaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatgggtc	attaaatcag	ttatggttcc	tttggtcgct	cgctcctctc	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgacggggcg	tgacccccct	9840
tcocgggggg	ggatgctgtc	atttatcaga	tcaaaaccaa	cccggtgagc	tcctccccgg	9900
ctocggcccg	gggtcggggc	ccggcggtct	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccg	tggcgggcgc	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgc	taccatgggt	accacgggtg	acggggaatc	aggggttcgat	tcgggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
cgacccgggg	aggtagtgac	gaaaaataac	aatacaggac	tccttcgagg	ccctgtaatt	10200
ggaatgagtc	cctttaacga	ggatccattg	ggatccattg	gagggcaagt	ctgggtgccg	10260
cagccgcgggt	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcgggcgggt	cgccgcgagg	cgagtcaacc	cccgcccccg	10380
ccccttgccct	ctcggcgccc	cctcgatgct	cctcgatgag	tgtcccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccca	gccgcctgga	taccgcagct	10500
aggaataatg	gaataggacc	gcggttctat	tttggtgggt	ttcggaactg	aggccatgat	10560
taagaggggac	ggccggggggc	attcgtattg	cgccgctaga	ggtgaaattc	ttggaccggc	10620

gcaagacgga	ccagagcgaa	agcatttgcc	aagaatggtt	tcattaatca	agaacgaaa	10680
tcggagggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcggcg	gcgttattcc	catgaccgcg	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggt	tgcaaagctg	aaacttaaa	gaattgacgg	aagggcacca	10860
ccaggagtg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccgaca	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tggttggtgg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggctg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgccggggc	11160
tgacgcgcg	ctacactgac	tggtcagcgc	tggtccctacc	ctgcgcggcg	aggcgccggg	11220
aacccggtga	acccattcg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccg	taagtgcggg	tcataagctt	gcgttgatta	agtcctgcc	ctttgtacac	11340
acccgctgc	gctactaccg	attggatggt	ttagtggagc	cctcggatcg	gccccgcgg	11400
ggtcggccca	cggccctggc	ggagcgctga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagt	gtaacagggt	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcgcgct	ggcccgctct	ccccgtcttg	tggtgtgtct	cgccgggagg	11580
cgcgtgcgtc	cgggtcccg	tcgcccgcgt	gtggagcgag	gtgtctggag	tgagggtgaga	11640
gaaggggtgg	gtggggtcgg	tctgggtccg	tctgggaccg	cctccgattt	ccccctcccc	11700
tccccctctc	ctcgtccggc	cttgacctcg	ccacctacc	gcggcgcgcg	ctgctcgccg	11760
cgcgtcttgc	tctttcccg	ccgctctctc	cgtgtctacg	aggggcggtg	cgctcgttacg	11820
ggtttttgac	ccgtcccggg	ggcgttcggg	cgtcggggcg	cgcgctttgc	tctcccggca	11880
cccatccccg	ccgcggctct	ggcttttcta	cgttggtctg	ggcggttgct	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcgcgg	tcccgatgcc	acgcttttct	ggcctcgcgt	12000
gtcctccccg	ctcctgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaagga	12060
ccccgggggg	gtcgccttgc	cgcccccagg	gtcggggggc	gggtggggccc	gtaggggaagt	12120
cggtcggttc	ggcggtctct	cctcagactc	catgaccctc	ctccccccgc	tgcccgcggt	12180
cccgagcgcg	cggtcggtg	gggggggtga	tgctggagc	ccccctgggc	gcccgtgggg	12240
cccgacccgc	gcccgcggct	tgcccgatct	ccgcgggtcg	gtcctgtcgg	tgccgggtcgt	12300
gggttcccgt	gtcgttccc	tggttttccg	ctcccgaacc	tttttttttc	ctccccccca	12360
cacgtgtctc	gtttctgttc	tgctggccgg	cctgaggcta	ccccctcggt	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tcgttggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	ctcacacccg	aaataccgat	acgactctta	gcgggtggatc	actcggtcgt	12540
tgctgcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttccctc	cggggctacg	cctgtctgag	12660
cgtcgggtga	cgatcaatcg	cgtcaccgcg	tgcggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	acccgggtcg	ggcctccgt	ctcccgaagt	tcagacgtgt	12780
ggcgcggttg	cgggtgtggc	cgcgcgcggc	cgtcgcggag	cctgggtctcc	cccgcgcatc	12840
cgcgtctcgc	gcttcttccc	gctccgcgct	tcccgccctc	gcccgtgcac	cccggctcgt	12900
gcctcgcgtc	ggcgccctcc	ggaccgctgc	ctcaccagtc	tttctcgggt	ccgtgccccg	12960
tgggaaacca	ccgcgcccc	gtggcgcccc	gggttgggcg	cgtccgcata	tgctctggtc	13020
aggttggcg	gttgaggtg	tgctgcgcgc	gaggttggtg	tcgggtccct	gcggccgcgg	13080
gggtgtcggg	gtggcggtcg	acgagggccg	gtcggtcgoc	tgcggtggtt	gtctgtgtgt	13140
gtttgggtct	tcgcgtgggg	gagggcggtg	gcaccgctcg	cgggggtggc	gcggtcgccc	13200
ggcgccgcgc	accctcggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatccccgg	tggtcgttgc	agggaggggt	tggtcgtccc	cgtccgtccg	tccctccctc	13320
cctcgggtgg	cgccttcgcg	ccgcacgcgg	ccgctaaggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cctcgtctcc	gcttctcctt	caccggggcg	gtaccgcgtc	cggcgccggc	13440
ccgcgggacg	cgcggcggtc	cgtgcgcgga	tgcgagtca	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gacccgttgc	gtcccggctt	ccctgggggg	gacccggcgt	13560
ctgtgggctg	tgctcccg	gggttgctg	tgagtaagat	cctccacccc	cgccgcccct	13620
ccctcccgcc	ggcctctcgg	ggacccccct	agacgggtcg	cgggtcgtc	ctcccggtgc	13680
gcccgggtgc	gtctcttttc	cgcccgccct	ctcgtctct	tcttcccgcg	gctgggcgcg	13740
tgtccccctt	ttctgaccgc	gacctcagat	cagacgtggc	gacccgctga	atttaagcat	13800
attagtcagc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaaacagg	13860
gaagagccca	gcgcgcaatc	cccgccgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gacccactcc	ccggcgccgc	tctggtgggg	cccaagtcc	tctgatcgag	gcccagcccc	13980
tggaagggtg	gaggccggtg	gcggcccccg	cgcgcggggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaaactcca	tctaaggcta	aataccggca	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagtgtga	agaactttg	aagagagagt	14160
tcaagagggc	gtgaaaccgt	taagaggtaa	acgggtgggg	tccgcgcagt	ccgccccggg	14220
gattcaaccc	ggcgggcgcg	gtccggccgt	gcccgggtgt	cccgccggat	ctttcccgct	14280
ccccgttcc	cccgacccct	ccacccgcgc	gtcgttcccc	tcttctctcc	cgcgtccggc	14340
gcctccggcg	gcgggcccgg	gggttggtgt	gggttggtgg	cgcggggcgg	gcgggggggt	14400
gggtcggcgg	gggaccgccc	ccggccggcg	accggccgcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tccgggaagg	ccgggaaggc	ccgggtggga	aggtgggtcg	14520
ggggggggcg	cgcgtctcag	ggcgcccgca	accacctcac	cccgagtggt	acagccctcc	14580
ggccgcgctt	tcgccgaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgtctctc	14640

ctctcccccc	gtccgcctcc	cgggcgggcg	tggggggtggg	ggccggggccg	ccctccccc	14700
ggcgcgaccg	ctctcccacc	ccctcccgctc	gcctctctcg	gggcccgggtg	ggggggcgggg	14760
cggactgtcc	ccagtgcgcc	ccggggcgctc	tcgcgcgcgtc	gggtcccgggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcacgaa	gccgagcgca	cgggggtcggc	ggcgatgtcg	gctaccacc	14880
cgaccgcgtc	tgaaacacgg	accaaggagt	ctaaccggtg	cgcgagttag	gggctcgcc	14940
gaaagccgcc	gtggcgcaat	gaagggtgaag	ggccccgccc	ggggggcccg	ggtgggatcc	15000
cgaggccctc	ccagtccgcc	gagggcgcac	caccggcccc	tctcgcccc	cgcgccgggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggagggtcc	gtagcgggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	atagggcgga	aagactaatc	gaaccatcta	gtagctggtt	ccctccgaag	15240
tttccctcag	gatagtgggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaaccta	tctcaaaactt	taaattgggt	15360
agaagcccg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgccctag	tgggccactt	15420
ttggtaagca	gaactggcgc	tgcgggatga	accgaacgcc	gggttaaagg	gcccgatgcc	15480
gacgctcatc	agaccccgaa	aaagggtgtg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccgcccgctc	gccgcagtcg	gaacggaacg	15660
ggacggggag	ggccgcgggt	gcccgcgtctc	cgggggtcggg	ggtgctgggc	ggggggcccg	15720
cccccgctc	ccctccgcgc	gcccgggttcg	cccccgcgcc	gtcggggccc	gcccggacct	15780
cgccgcgacg	agtaggagg	ccgctgcgggt	gagccttgaa	gcctaggggc	cgggcccggg	15840
tggagccgcc	gcaggtgcag	atcttggtgg	tagtagcaaa	tattcaaacg	agaactttga	15900
aggccgaagt	ggagaagggt	tccatgtgaa	cagcagttga	acatgggtca	gtcgggtcctg	15960
agagatgggc	gagtgccgtt	ccgaagggac	gggcagttggc	ctccgttgcc	ctcgccgat	16020
cgaaggggag	tccgggttcag	atccccgaat	ccggagtggtc	ggagatgggc	gccgcgaggc	16080
cagtgccgta	acgcgaccga	tcccggagaa	gcccggcgga	ggcctcgggg	agagtctctc	16140
ttcttttgtg	aagggcaggg	cgccctggaa	tgggttcgcc	ccgagagagg	ggcccggtgc	16200
ttggaaagcg	tcgcggttcc	ggcgccggtcc	ggtgagctct	cgctggccct	tgaaaatccg	16260
ggggagaggg	tgtaaatctc	gcccgcgggcc	gtacccatat	ccgcagcagg	tctccaaggt	16320
gaacagcctc	tggcatgttg	gaacaatgta	ggtaaaggaa	gtcggcaagc	cggatccgta	16380
acttcgggat	aaggattggc	tctaagggtc	gggtcggtcg	ggctggggcg	cgaagcgggg	16440
ctgggcgcgc	gcccgcggctg	gacgagggcg	cgccgcctct	tcccacgtcc	ggggagaccc	16500
cccgtccttt	ccgcccgggc	ccgccctccc	ctcttccccg	cggggccccc	tcgtcccccg	16560
cgctcgcgc	acgtctcttc	ccctctcttc	cttcccgctc	gggggcgggt	cgggggtcgg	16620
cgcgcgccgc	gggctccggg	gcccgcgggtc	caaccccgcg	ggggttcggg	agcgggagga	16680
accagcggtc	cccggtgggg	cgggggggccc	ggacactcgg	ggggccggcg	gcccggcgga	16740
ctctggacgc	gagccggggc	cttcccgtgg	atcgccctcag	ctgcggcggg	cgctcgccgc	16800
gctcccgggg	agcccggcg	gtgcccggcg	gggtccccctc	cccgcggggc	ctcgctccac	16860
ccccccatcg	cctctcccga	ggtgctgggc	ggggcggggc	gggcgtgtcc	cgcgcggtg	16920
gggggaaact	ccgcgtcggt	gttccccgcg	cgggtccgcg	ccccggggcg	cggttttccg	16980
cgcgcccgcc	ccgcctcggc	cgcgccctag	cagccgactt	agaactggtg	cggaccagg	17040
gaatccgact	gtttaattaa	aacaaagcat	cgcgaaaggc	cgcgggcggt	ggtgacgcga	17100
tgtgatttct	gcccagtgct	ctgaatgtca	aagtgaagaa	attcaatgaa	gcccgggtaa	17160
acggcgggag	taactatgac	tctcttaagg	tagccaaatg	cctcgtcatc	taattatgta	17220
cgcgcataga	tggatgaacg	agattcccac	tgtccctacc	tactatccag	cgaaccaca	17280
gccaagggaa	cggtctgggc	ggaatcagcg	gggaaagaa	accctgttga	gcttgactct	17340
agtctggcac	ggtgaagaga	catgagaggt	gtagaataag	tgggaggccc	ccggcgcccc	17400
gcccgcgtcc	cgcggtcggg	tcggggcacg	ccggccctcgc	gggcccggcg	tgaaatacca	17460
ctactctcat	cgttttttca	ctgaccgggt	gaggcggggg	ggcgagcccc	gaggggctct	17520
cgcttctggc	gccaaagcgtc	cgccccgcgc	gtgcggggcg	gcccagcccc	ctccggggac	17580
agtgccaggt	ggggagtttg	actggggcg	tacacctgtc	aaacggtaac	gcaggtgtcc	17640
taaggcgagc	tcaggggagga	cagaaacctc	ccgtggagca	gaagggcaaa	agctcgcttg	17700
atcttgattt	tcagtacgaa	tacagaccgt	gaaagcgggg	cctcacgatc	cttctgacct	17760
tttgggtttt	aagcaggagg	tgtcagaaaa	gttaccacag	ggataaactgg	cttgtggcgg	17820
ccaagcgttc	atagcgacgt	cgctttttga	tccttcgatg	tcggctcttc	ctatcattgt	17880
gaagcagaat	tcaccaagcg	ttggattgtt	caccactaa	tagggaacgt	gagctgggtt	17940
tagaccgtcg	tgagacaggt	tagttttacc	ctactgatga	tgtgtgtgtg	ccatggtaat	18000
cctgctcagt	acgagaggaa	ccgcaggttc	agacatttgg	tgtatgtgct	tggctgagga	18060
gccaatgggt	cgaagctacc	atctgtggga	ttatgactga	acgcctctaa	gtcagaatcc	18120
gccaagcgcg	aacgatacgg	cagcgccgaa	ggagccctcg	ttggccccgg	atagccgggt	18180
ccccgctcgt	ccgcctcggc	gggggtcccc	cgctcgcccc	cgcgcccgcg	gggtctcccc	18240
ccgcggggcg	tcgggacggg	ggtcdggtgc	ggagagccgt	tcgtcttggg	aaacgggggtg	18300
cgcccggaaa	ggggcccgcc	ctctcgcccc	tcacgttgaa	cgacgttcg	tgtggaacct	18360
ggcgctaaac	cattcgtaga	cgacctgctt	ctgggtcggg	gtttcgtagc	tagcagagca	18420
gctccctcgc	tgcatctat	tgaaagttag	ccctcgacac	aagggtttgt	ctctcggggc	18480
tttcccgctc	cagcccgctc	cgctcgacgc	cgaccgtgtc	gcccgggggg	cgtcacgggg	18540
gcggtcgcc	cgcccccgcg	gcggttgccc	gaacgaccgt	gtggtggttg	ggggggggat	18600
cgtcttctcc	tcgctctccc	gaggacgggt	cgtttctctt	tcccttcccg	tcgctctcct	18660

tgggtgtggg	agcctcgtgc	cgtcgcgacc	gcggcctgcc	gtcgcctgcc	gccgcagccc	18720
cttgccctcc	ggccttggcc	aagccggagg	gcggaggagg	gggatcggcg	gcggggcgga	18780
ccgcgccgcg	gtgacgcacg	gtgggatccc	catcctcggc	gcgtccgtcg	gggacggccg	18840
gttgaggagg	cgaggagggt	ttttcccggt	aacgcgcggt	tcggcgccag	gcctctggcg	18900
gccggggggg	cgctctctcc	gccccgagcat	ccccactccc	gccccctctc	ttcgcgcgcc	18960
gcggcgccgc	cgtgcgtacg	aggggaggat	gtcgcgggtg	ggaggcggag	aggggtccggc	19020
actttgtttt	ttccattttt	tcccccccaa	cttcggagggt	cgaccagtac	tccggggcgac	19080
cccccccccc	ttttttttcc	cccgatgctg	gaggtcgacc	agatgtccga	aaagtgtcccc	19140
tttttttttt	tttggtgtgc	cggagcggcg	gggccaactct	ggactctttt	tttttttttt	19200
catatagggtc	gaccagttact	gggaaccttta	gttcgaccag	ttgtccgtct	tttactcctt	19260
gaccagatat	ccgaaagtcc	ccgggtggta	ctttgtcttt	ttctgaaaat	cccagagggtc	19320
tttttttttt	tttggtgtgc	tctctttccc	tttactcttc	cccacagcga	ttctcttttt	19380
atatacttat	aggaggagggt	cgaccagtac	cttatataca	tgtaaatagt	gtgtacgttt	19440
tccaccgatg	atggagggtcg	accagatgtc	tccgggcgac	actttgtttt	tttttttttt	19500
ccgcgacgcg	gcgggctcac	tctggactct	cgaaagtgtc	ccgtcccccc	cctccccccc	19560
tggaacctta	aggtcgacca	gttgctcgctc	tttttttttt	tttttttttt	tttaaaatttc	19620
tactttgtct	ttttctgaaa	atcgcagagg	tttactcat	tcatatagggt	cgaccgggtgg	19680
ataaattatc	tgatctagat	ttgtttttct	tcgaccagat	gtcagaaagt	ctgggtgggtcg	19740
tttgtgttgt	tttgttttgt	tttgttttgt	gtttttcagt	tttgtgttgt	tttgtgttgt	19800
tttgtgttgt	gttgtgttgt	gttgtgttgt	gtttgttgtg	gttgtgttgt	gttgtgttgt	19860
gttgtgttgt	gttgtgttgt	gttgtgttgt	tttgtgttgt	tttgtgttgt	tttgtgttgt	19920
ttgtttgtctg	ttgtttgtgt	tttgtcggtg	cgaaacagttg	tccctaaccg	agttttttttg	19980
tacacaaaca	tgacttttgt	tttaaaataaa	tttttaaaat	aaatgcgaaa	atcgaccaat	20040
tatccctttc	cttctctctc	tttttaaaaa	attttctttg	tggtgtgtgtg	tggtgtgtgtg	20100
tggtgtgtgtg	tgctgtgtgtg	tggtgtgtgtg	cgtgcagcgt	gcgcgcgcctc	gttttataaaa	20160
tacttataat	aataggctcg	cggtgtgtgtg	tagcttcccc	gactccagag	gttttataaaa	20220
gcagacttct	gagttcgagg	ccagcctgggt	ctacagagga	accctgtctc	gaaaaatgaa	20280
aataaataca	tacatacata	catacataca	tacatacata	catacataca	tacatatgag	20340
gttgaccagt	tgctcaatcct	ttagaatttt	aatgtgatag	aatgtgatag	agagatagat	20400
aatagataga	tgatagagat	gatacaaaata	taggtttttt	tttcagtaaa	tatgagggtg	20460
attaaccact	tttccctttt	taggtttttt	tttttttccc	ctgtccatgt	ggttgctggg	20520
atttgaactc	aggaccctgg	cagggtcaact	ggaaaacgtg	ttttctatat	atataaatag	20580
tggtctgtct	gtgtgtgtgt	tggtgtgtgt	ctgtctgtct	tgctgtgtgt	cttgccttgc	20640
tgcttttttt	tttcttctga	gacagtattt	ctctgtgtaa	cctgggtgcc	tgaaactcac	20700
tctgtagacc	agcctggcct	caatcgaaact	cagaaatcct	cctgcctctt	gtctacctcc	20760
caattttgga	gtaaaagggtg	gtacacacac	gcctggcat	tattatcatt	atcattatta	20820
attttattat	tagacagaac	gaaatcaact	agttggctct	gtttcggttaa	ttcatttgaa	20880
attagttgga	ccaattagtt	ggctgggtttg	ggaggtttct	tttggtttccg	atttgggtgt	20940
ttgtgggggt	ggggatcagg	tatctcaacg	gaatgcata	agggttaagg	gagatggctc	21000
gatttttgta	aagattactt	ttcttagtct	gaggaaaaaa	taaaataata	ttgggtctacg	21060
tttcatttgt	tcatttctat	ttctctttct	ttctttcttt	ctttcagata	aggaggtcgg	21120
ccagttcctc	ctgccttctg	gaagatgtag	gcattgcatt	gggaaaagca	ttgtttgaga	21180
gatgtgctag	tgaaccagag	agtgtggatg	tcaagccgta	taatgtttat	tacaatatag	21240
aaaagtcta	acaaagtgat	cttttaactt	tttttttttt	tttctccttc	tacttctact	21300
tggtctcact	ctgccaccaa	cgcgctttgt	acattgaatg	tgagctttgt	tttgcttaac	21360
agacataat	tttttctttt	ggttttgtct	gacatgggtt	ccctttctat	ccgtgcaggg	21420
ttcccagacg	gccttttgag	aataaaatgg	gagggcagaa	ccaaagtctt	ttgaataaag	21480
caccacaact	ctaacctgtt	tggtgtttt	ccttcccagg	gcacagatct	ttcccagcat	21540
ggaaaagcat	gtagcagttg	taggacacac	tagacgagag	caccagatct	cattgtgggt	21600
gggtgtgaac	caccacccat	gtggttgctt	gggatttgaa	ctcaggatct	tcagaagacg	21660
agtcagggtc	ctaaaccgat	gagccatctc	tccagccctc	ctacattcct	tcttaaggca	21720
tgaatgatcc	cagcatggga	agacagtctg	ccctctttgt	ggatatcac	catatactca	21780
ataaaataat	gaaatgaaatg	aagtctccac	gtatttattt	cttcgagcta	tctaaaattct	21840
ctcacagcac	ctcccctcc	cccacactgc	ctttctccct	atgtttgggt	ggggctgggg	21900
gaggggtggg	gtgggggcag	ggatctgcat	gtcttcttgc	aggtctgtga	actatttgcg	21960
atggcctggt	tctctgaact	gttaggcctt	gtctatccag	aggtgactg	gctagttttc	22020
tacctgaagt	ccctgagtg	tgatttccct	gtgaattc			22080
						22118

<210> 19
 <211> 175
 <212> DNA
 <213> Mus musculus

<400> 19	ctccgcgcg	gccccgtgt	tcgcccgttc	cgtggcgcg	acaatgcggt	tggtcgtcca	60
	cgtgtgcgtg	tccgtgcagt	gcggtgtgtg	agtgccctcg	tctcctctc	ctccccggca	120

gcgttcccac ggttggggac caccggtgac ctgcacctct tcgggcctgg atccg

175

<210> 20
<211> 755
<212> DNA
<213> Mus musculus

<400> 20							
ggtctggtgg	gaattgttga	cctcgtctct	gggtgcgggc	tttggggaac	ggcgggggtcg		60
gtcgtgcccc	gcgcgggacg	tgtgtcgggg	cccacttccc	gctcagagggt	ggcgggtggcg		120
gcggcggttg	tagtctcccc	tggtgcgctct	tcccgggctc	ttgggggggg	tgccgtcggt		180
ttcggggccg	gcgttgcttg	gcttacgcag	gcttggtttg	ggactgcctc	aggagtcgtg		240
ggcgggtgtga	ttcccgcggg	ttttgcctcg	cgtctgcctg	ctttgcctcg	ggtttgcttg		300
gttcgtgtct	cgggagcggg	ggtttttttt	tttttcgggt	cccggggaga	gggggtttttc		360
cgggggacgt	tcccgctgcc	ccctgcgcgc	ggtgggtttt	cgtttcgggc	tgtgttcgtt		420
tccccttccc	cgtttcgccg	tcggttctcc	ccggtcggtc	ggccctctcc	ccggtcggtc		480
gcccggccgt	gctgcgcggc	cccccttct	gggggggatg	cccgggcacg	cacgcgtccg		540
ggcggccact	gtggtccggg	agctgctcgg	caggcgggtg	agccagttag	aggggctgca		600
tgcccccgcg	ggctcccgtg	gccgacgcgg	cgtgttcttt	ggggggggct	gtgcgtgcgg		660
gaaggctcgc	cacgttgtcg	gtccttgcca	gggaaagagg	cttttttttt	ttaggggggtc		720
gtccttcgtc	gtcccgtcgg	cggtggtatc	ggcct				755

<210> 21
<211> 463
<212> DNA
<213> Mus musculus

<400> 21							
ggcgcaggtg	cgtctgcggg	ttgggggtcg	tccggccccg	tcgtcctccg	ggaaggcgtt		60
tagcgggtac	cgtcgccgcg	ccgaggtggg	cgcacgtcgg	tgagataacc	ccgagcgtgt		120
ttctggttgt	tgccggcggg	ggctccgggtc	gatgtcttcc	cttccccctc	tccccgagggc		180
caggtcagcc	tccgcctgtg	ggcttcgtcg	gcctctctcc	ccccctcac	gtccctcgcg		240
agcagagccc	tccgttcgac	cttccttccg	ccttcccccc	atctttccgc	gtcccggttg		300
ccccgggggt	ttcacggcgc	ccccacgct	cctccgcctc	tccgcccgtg	gtttggacgc		360
ctgggtcccg	tctccccgcc	aaaccccggt	tgggttggtc	tccggccccg	gcttgcctct		420
cgggtctccc	aacccccggc	cggaagggtt	cgggggttcc	ggg			463

<210> 22
<211> 378
<212> DNA
<213> Mus musculus

<400> 22							
ggattcttca	ggattgaaac	ccaaaccggg	tcagtttctt	ttccgggtcc	ggcggggggg		60
ggcggccccg	ggcgggtttg	tgagttagat	aacctcgggc	cgatcgacag	ccccccgtgg		120
cggcgacgac	ccattcgaac	gtctgcccta	tcaactttcg	atggtagtcg	atgtgcctac		180
catggtgacc	acgggtgacg	gggaatcagg	gttcgattcc	ggagagggag	cctgagaaac		240
ggctaccaca	tccaaggaag	gcagcaggcg	cgcaaattac	ccactcccga	cccggggagg		300
tagtgacgaa	aaataacaat	acaggactct	ttcgaggccc	tgtaattgga	atgagtcac		360
tttaaatcct	ttaagcag						378

<210> 23
<211> 378
<212> DNA
<213> Mus musculus

<400> 23							
gatccatttg	agggcaagtc	tggtgccagc	agccgcggta	attccagctc	caatagcgta		60
tattaaagtt	gctgcagtta	aaaagctcgt	agttggatct	tgggagcggg	cgggcgggtcc		120
gccgcgaggg	gagtcaccgc	ccgtccccgc	cccttgccctc	tcggcgcccc	ctcgatgctc		180
ttagctgagt	tgtcccgcgg	ggcccgaagc	gtttactttg	aaaaaattag	agttgtttca		240
aagcaggccc	gagccgcctg	gataccgcca	gctaggaaat	aatggaatag	gaccgcgggtt		300
cctattttgt	ttggttttgc	gaactgagcc	catgattaag	ggaaacggcc	gggggcatte		360
ccttattgcg	cccccta						378

<210> 24
<211> 719

<212> DNA

<213> Mus musculus

<400> 24

ggatctttcc	cgtccccgt	tctccccgc	ccctccacc	gcgcgtctcc	cccttcttt	60
tccctctcc	ggagggggg	gaggtgggg	cgctggggc	gggtcgggg	tggggtcgg	120
gggggaccg	ccccggccg	caaaaggcc	ccgcggggc	cacttcaacc	gtagcgggt	180
gcccgcacc	gctacgagc	ggctgggaag	gcccgcggg	gaatgtggc	cgggggggg	240
ggcgcgtct	agggcgccg	gaaccacct	accccgagt	ttacagccct	cggcccgcc	300
tttcgcggaa	tccccgggg	gaggggaag	ccgataccg	tgcgcgcgt	tttccctcc	360
ccccgtccg	ctcccgggc	ggcgtgggg	tggggggcg	gcccgcctc	ccacgcccgt	420
ggtttctct	tctcccggt	tccgcccgt	tgggggggg	agcccgggt	ggggcgggg	480
ggactgtcc	cagtgcgcc	cgggcgtcg	cgcccgctc	ggcccgggg	gttctctcg	540
tcacgcgcc	cccgcgaag	ccgagcgac	ggggtcggc	gcgatgtcg	ctaccaccc	600
gacccgtct	gaaacacga	ccaaggagt	taacgcgtc	gcgagtcag	ggctcgcac	660
aaagccgcc	tggcgcaat	aaggtgaag	gccccgtcc	ggggcccgag	gtgggatcc	719

<210> 25

<211> 685

<212> DNA

<213> Mus musculus

<400> 25

cgaggcctct	ccagtcggc	gagggcgac	caccggccc	tctcgccgc	cgctcgggg	60
aggtggagca	cgagcgtac	cgtaggacc	cgaaagatg	tgaactatg	ctgggcaggg	120
cgaagccaga	ggaaactct	gtggaggtc	gtagcggtc	tgacgtgca	atcggtcgt	180
cgacctgggt	atagggcgga	aagactaat	gaaccatct	gtagctgggt	ccctccgaag	240
tttccctcag	gatagctgg	gctctcgaa	ccttcggaag	cagttttat	cgggtaaagg	300
cggaatggat	taggaggtct	tggggccgga	aacgatctca	aactatttt	caaacttta	360
atgggtaagg	aagcccggc	cgctggcgt	gagccgggc	tggaatgca	gtgcctagt	420
ggccactttt	ggtaagcaga	actggcgct	cgggatgaac	cgaacgccg	gttaaggcgc	480
ccgatgccga	cgctcatcag	accccgaaa	aggtgttgg	tgatatagac	agcaggacgg	540
tggccatgga	agtcggaat	cgctaaggag	tgtgtaacaa	ctcacctgcc	gaatcaacta	600
gccctgaaaa	tggatggcg	tggagcgtc	ggcccatacc	cggccgtcg	cggcagtcgg	660
aacgggacgg	gacgggagcg	gccgc				685

<210> 26

<211> 5162

<212> DNA

<213> Artificial Sequence

<220>

<223> Chimeric bacterial plasmid

<400> 26

gacggatcgg	gagatctccc	gatcccctat	ggtcgactct	cagtacaatc	tgctctgatg	60
ccgcatagtt	aagccagtat	ctgctccctg	cttgtgtgtt	ggaggtcgct	gagtagtgcg	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgaccga	caattgcatg	aagaatctgc	180
ttagggttag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccggc	tggctgaccg	cccaacgacc	360
cccgcccat	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggg	aaactgccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacggcc	octattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	ccatttgacg	caaattggcg	780
gtaggcgtgt	acggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
tgctttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttggtacc	900
gagctcggat	cgatatctgc	ggccgcgtcg	acgggaattca	gtggatccac	tagtaacggc	960
cgccagtggt	ttcggtgtct	ttcggtgtct	gcgagggcca	gctgttgggg	tgagtactcc	1020
ctctcaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaa	cgaggaggat	1080
ttgatattca	cctggcccgc	ggtgatgcct	ttgaggggtg	ccgcgtccat	ctgggtcagaa	1140
aagacaatct	ttttgttgtc	aagcttgagg	tgtggcaggc	ttgagatctg	gccatacact	1200
tgagtgacaa	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	caggtccaac	1260
tgcaggtcga	gcatgcatct	agggcgggcca	attccgcccc	tctccctccc	ccccccctaa	1320

cggtactggc	cgaagccgct	tggaataaagg	ccggtgtgcg	tttgtctata	tgtgattttc	1380
caccatattg	ccgtctttttg	gcaatgtgag	ggcccggaaa	cctggccctg	tcttcttgac	1440
gagcattcct	aggggtctttt	ccccctctcg	caaaggaatg	caaggtctgt	tgaatgtcgt	1500
gaagggaagca	gttcctcttg	aagcttcttg	aagacaaaca	acgtctgtag	cgaccctttg	1560
caggcagcgg	aacccccccac	ctggcgacag	gtgcctctgc	ggccaaaagc	cacgtgtata	1620
agatacacct	gcaaaggcgg	cacaacccca	gtgccacgtt	gtgagttgga	tagttgtgga	1680
aagagtcaaa	tggctctcct	caagcgtatt	caacaagggg	ctgaaggatg	cccagaagg	1740
accccatgtg	atgggatctg	atctggggcc	tcgggtgcaca	tgctttacat	gtgttttagtc	1800
gaggttaaaa	aaacgtctag	gcccccgaa	ccacggggac	gtgggttttc	tttgaaaaac	1860
acgatgataa	gcttgccaca	acccgggagc	caccgtctgc	caccatgggtg	agcaaggggcg	1920
aggagctggt	caccgggggtg	gtgcccaccc	tggtcgagct	ggacggcgac	gtaaaacggcc	1980
acaagttcag	cgtgtccggc	gagggcgagg	gcgagtcacc	ctacggcaag	ctgaccctga	2040
agttcatctg	caccacccgg	aagctgcccc	tgccctggcc	caccctcgtg	accaccctga	2100
cctacggcgt	gcagtgcttc	agccgctacc	ccgaccacat	gaagcagcac	gacttcttca	2160
agtccgccat	ccccgaaggc	tacgtccagg	agcgcacccat	cttcttcaag	gacgacggca	2220
actacaagac	ccgcgcggag	gtgaagttcg	aggcgacac	cctgggtgaa	cgcatcgagc	2280
tgaagggcat	cgacttcaag	gaggacggca	acatcctggg	gcacaagctg	gagtacaact	2340
acaacagcca	caacgtctat	atcatggccg	acaagcagaa	gaacggcatc	aagggtgaact	2400
tcaagatccg	ccacaacatc	gaggacggca	gcgtgcagct	cgccgaccac	taccagcaga	2460
acacccccat	ccgcgcggc	cccgctgctg	tgcccgaaca	ccactacctg	agcaccaggt	2520
ccgccttgag	caaagacccc	aacgagaagc	gcgatcacat	ggctcctgctg	gagttcgtga	2580
ccgcgcggcg	gatcactctc	ggcatggagc	agctgtacaa	gtaaagcgcc	cctagagctc	2640
gctgatcagc	ctcgactgtg	cctctagtgt	ccagccatct	gttgtttgcc	cctccccctg	2700
gccttctctg	accctggaag	gtgccactcc	cactgtcctt	tcctaataaa	atgaggaaat	2760
tgcacgcat	tgtctgagta	ggtgtcattc	tattctgggg	gggtggggtg	ggcaggacag	2820
caagggggag	gattgggaag	acaatagcag	gcagctggg	gatgcgggtg	gctctatggc	2880
ttctgagcgg	gaaagaacca	gctggggctc	gagtgatttc	tagttgtggt	ttgtccaaac	2940
tcatcaatgt	atcttatcat	gtctgtatac	cgctgacctc	tagctagagc	ttggcgtaat	3000
catggtcata	gctgtttcct	gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	3060
gagccggaag	cataaagtgt	aaagcctggg	gtgcctaata	agttagctaa	ctcacattaa	3120
ttgcgttgcg	ctcactgccc	gctttccagt	cggaacacct	gtcgtgccag	ctgcattaat	3180
gaatcgccca	acgcgcgggg	agaggcggtt	tgcgatttgg	gcgctcttcc	gcttccctgc	3240
tcactgactc	gctgcgctcg	gtcggtcggc	tgccggcagc	ggtatcagct	cactcaagg	3300
cggtaatcac	gttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	3360
gccagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcggttttc	cataggctcc	3420
gccccctga	cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aaccgcagac	3480
gactataaag	atagatcagg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	3540
ccctgcgcgt	taccggatac	ctgtccgcct	ttctcccttc	gggaagcgctg	gcgctttctc	3600
aatgctcacg	ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	3660
tgcacgaacc	ccccgttcag	cccgaccgct	gcgccttacc	cggttaactat	cgtcttgagt	3720
ccaaccgggt	aagcacgcag	ttatcgccac	tgccagcagc	cactggtaac	aggattagca	3780
gagcgaggta	tgtaggcggt	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	3840
ctagaaggac	agtatttggg	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	3900
ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	cggtgggttt	tttgtttgca	3960
agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	4020
ggtctgacgc	tcagtggaa	gaaaactcac	gttaagggat	tttggtcagt	agattatcaa	4080
aaaggatcct	cacctagatc	cttttaaaat	aaaaatgaag	ttttaaatca	atctaagta	4140
tatatgagta	aacttgggtc	gacagttacc	aatgcttaat	cagtgaggca	cctatctcag	4200
cgatctgtct	atcttcttca	tccatagttg	cctgactccc	cgctcgttag	ataactacga	4260
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	4320
cggtccaga	tttatcagca	ataaaccagc	cagccggaa	ggccgagcgc	agaagtggtc	4380
ctgcaacttt	atccgcctcc	atccagttca	ttaatgtgtg	ccgggaagct	agagtaagta	4440
gttcgcccag	taatagtttg	cgcaacggtg	ttgccattgc	tacaggcatc	gtgggtgtcac	4500
gctcgtcgtt	tggtatggct	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	4560
gatcccccat	gttgtgcata	aaagcgggta	gctccttcgg	tcctccgac	gttgtcagaa	4620
gtaagttggc	cgcagtggtt	tactcatggg	ttatggcagc	actgcataat	tctcttactg	4680
tcatgccatc	cgtaagatgc	ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	4740
aatagtgat	gcggcgaccg	agttgtctct	gcccggcgct	aatacgggat	aataccggc	4800
cacatagcag	aactttaaaa	gtgctcatca	ttggaaaacg	ttcttcgggg	cgaaaactct	4860
caaggatcct	accgctgttg	agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	4920
cttcagcatc	ttttactttc	accagcggtt	ctgggtgagc	aaaaacagga	aggcaaaatg	4980
ccgcaaaaaa	gggaataaag	gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	5040
aatattattg	aagcatttat	caggggtatt	gtctcatgag	cggatacata	tttgaatgta	5100
tttagaaaaa	taaacaaata	gggggtccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	5160
tc						5162

<211> 5627
<212> DNA
<213> Artificial Sequence

<220>
<223> pMG plasmid from InvivoGen; IRES sequence modified
EMCV nucleotides 2736-3308

```

<400> 27
caccggcgaa ggaggcctag atctatcgat tgtacagcta gctcgacatg ataagataca      60
ttgatgagtt tggacaaacc acaactagaa tgcagtgaaa aaaatgcttt atttgtgaaa      120
tttgtgatgc tattgcttta ttgtgaaat ttgttgatgct attgctttat ttgtaaccat      180
tataagctgc aataaacaag ttaacaacaa caattgcatt cattttatgt ttcagggttca      240
gggggaggtg tgggaggttt tttaaagcaa gtaaaacctc tacaaatgtg gtagatccat      300
ttaaatgtta attaagaaca tgtgagcaaa agggccagcaa aaggccagga accgtaaaaa      360
ggcgcggttg ctggcggttt tccataggct ccgccccctt gacgagcatc acaaaaaatcg      420
acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttcccccc      480
tggaagctcc ctctgctgct ctctgttccc gacctgcccg cttaccggat acctgtcccg      540
ctttctccct tggggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc      600
gggttaggtc gttcgctcca agctgggctg tgtgcacgaa cccccgttc agccccgaccg      660
ctgctgcctt tccggttaact atcgtcttga gtccaacctg gtaagacacg acttatcgcc      720
actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga      780
gttcttgaag ttgtggccta actacggcta cactagaaga acagtatttg gtatctgcgc      840
tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac      900
caccgctggt agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg      960
atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaaactc      1020
acgttaaggg attttggtca tggctagtta attaagctgc aataaacaat cattattttc      1080
attggatctg tgtgttggtt ttttgtgtgg gcttggggga gggggaggcc agaactgactc      1140
caagagctac aggaaggcag gtcagagacc ccactggaca aacagtggct ggactctgca      1200
ccataacaca caatcaacag gggagtggag tggatcgagc tagagtcctg tacataactt      1260
acggtaaatg gccgcctggt ctgaccgccc aacgaccccc gccattgac gtcaataatg      1320
acgtatgttc ccatagtaac gccaataggg actttccatt gacgtcaatg ggtggagtat      1380
ttacggtaaa ctgcccactt ggcagtagta caagtgtatc atatgccaaag tacgccccct      1440
attgacgtca atgacggtaa atggcccgc tggcattatg cccagtacat gaccttatgg      1500
gactttccta cttggcagta catctacgta ttagtcatcg ctattaccat ggtgatgcgg      1560
ttttggcagt acatcaatgg gcgtggatag cggtttgact cagggggatt tccaagtctc      1620
caccctattg acgtccaatg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa      1680
tgtcgtaaca actccgcccc attgacgcaa atgggcggtg ggcgtgtacg gtgggaggtc      1740
tatataagca gagctcggtt agtgaaccgt cagatcgcc tggagacgcca tccacgtgtg      1800
tttgacctcc atagaagaca ccgggaccga tccagctccc gcggccggga acggtgcatt      1860
ggaacgcgga ttccccgtgc caagagtgac gtaagtaccg cctatagagt ctataggccc      1920
acccccctgg cttcttatgc atgctatact gtttttggct tgggggtctat acacccccgc      1980
tgacctctgt tataggtgat ggtatagctt agcctatagg tgtgggttat tgaccattat      2040
cacaactctc tttattggct atatgccaat acactgtcct tccataacat ggctctttgc      2100
tgtattttta caggatgggg tctcatttat acactgtcct tccagagactg acacggactc      2160
gtccccagtg ccgcaggttt ttattaaaca taacgtggga tctccacgcg aatctcgggt      2220
acgtgttccg gacatgggct cttctccggt agcggcgagg cttctacatc cgagccctgc      2280
tcccatgcct ccagegactc atggtcgctc ggcagctcct tgctoctaac agtggaggcc      2340
agacttaggc acagcacgat gcccaccacc accagtgtgc cgcacaaggc cgtggcggtg      2400
gggtatgtgt ctgaaaatga gctcggggag accagcttgc ccgctgacgc atttggaaga      2460
cttaaggcag cggcagaaga agatgcaggc agctgagttg ttgtgttctg ataagagtca      2520
gaggttaactc ccgttgcggt gctgttaacg gtggagggca gtgtagtctg agcagtactc      2580
gttgctgccc cgcgcgccac cagacataat agctgacaga ctaacagact gttcctttcc      2640
atgggtcttt tctgcagtca cccgggggat ccttcgaacg tagctctaga ttgagtcgac      2700
gttactggcc gaagcgcgtt ggaataaggc cggtgtgcgt ttgtctatat gttattttcc      2760
accatattgc cgtcttttgg caatgtgagg gcccggaaac ctggccctgt ctgtgtcgag      2820
agcatctcta ggggtctttc cctctcgcc aaaggaatgc aaggtctgtt gaatgtcggt      2880
aaggaagcag ttctcttgga agcttcttga agacaaacaa cgtctgtagc gaccttttgc      2940
aggcagcgga cccccacc accaggcggt tggcgacagg tgccctctgc gccaaaagcc acgtgtataa      3000
gatacacctg acaaccccag aacgctatct tctggggcct cggtgcacat gctttacatg      3060
agagtcaaat ggctctcttc aagcgtatct tctggggcct cagggggacg gggcgcctagt      3120
ccccattgta aacgtctagg gatattctact agttgtgacc ggcgcctagt gtgacaatt      3180
cgataatacc atgggtaagt ggcatagtat aatacgactc actataggag ggccaccatg      3240
aatcatcggc atagtatatc gctgagatca ccggtaggag ggccatcatg      3300
tcgactacta accttcttct ctttctctaca gctgagatca ccggtaggag ggccatcatg      3360

```

aaaaagcctg	aactcaccgc	gacgtctgtc	gcgaagtttc	tgatcgaaaa	gttcgacagc	3540
gtctccgacc	tgatgcagct	ctcggagggc	gaagaatctc	gtgcttttcag	cttcgatgta	3600
ggaggggcgtg	gatatgtcct	gcggggtaaat	agctgcgcgc	atgggtttcta	caaagatcgt	3660
tatgtttatc	ggcactttgc	atcggccgcg	ctcccgcattc	cggaaagtgc	tgacattggg	3720
gaattcagcg	agagcctgac	ctattgcac	tcccgcgcgtg	cacagggtgt	cacgttgcaa	3780
gacctgcctg	aaaccgaact	gcccgcgtgt	ctgcaaccgc	tgcgggagct	catggatgcg	3840
atcgtctgcgg	ccgatcttag	ccagacgagc	gggttcggcc	cattcggacc	gcaaggaatc	3900
ggtcaataca	ctacatggcg	tgatttcata	tgcgcgattg	ctgatcccca	tgtgtatcac	3960
tggcaaaactg	tgatggacga	caccgtcagt	gcgtccgtcg	cgcaggctct	cgtatgagctg	4020
atgctttggg	ccgaggactg	ccccgaagtc	cggcacctcg	tgacgcggga	tttcggctcc	4080
aacaatgtcc	tgacggacaa	tggccgcata	acagcggta	ttgactggag	cgaggcgatg	4140
ttcggggatt	cccaatacga	ggtcgccaac	atcttctct	ggaggccgtg	gttggcttgt	4200
atggagcagc	agacgcgcta	cttcgagcgg	aggcatccgg	agcttgcagg	atcgcgcggg	4260
ctccgggcgt	atatgtcccg	cattggctct	gaccaactct	atcagagctt	ggttgacggc	4320
aatttcgatg	atgcagcttg	ggcgacggg	cgatgcgacg	caatcgctcg	atccggagcc	4380
gggactgtcg	ggcgtaacac	aatcgccgcg	agaagcgcgg	ccgtctggac	cgtatggctgt	4440
gtagaagtac	tgcgcgatag	tggaaaccga	cgccccagca	ctcgtccgag	ggcaaaggaa	4500
tgagtgcaga	attcgctaga	gggccctatt	ctatagtgtc	acctaataatg	tagagctcgc	4560
tgatcagcct	cgactgtgcc	ttctagtgtc	cagccatctg	ttgtttgccc	ctcccccggtg	4620
ccttccttga	ccctggaagg	tgccactccc	actgtccttt	cctaataaaa	tgaggaaatt	4680
gcatcgcat	gtctgagtag	gtgtcattct	attctggggg	gtgggggtggg	gcaggacagc	4740
aaggggggagg	attgggaaga	caatagcagg	catgcgcagg	gccaattgc	tcgagcggcc	4800
gcaataaaaat	atctttatatt	tcattacatc	tgtgtgttgg	ttttttgtgt	gaatcgtaac	4860
taacatacgc	tctccatcaa	aacaaaacga	aacaaaacaa	actagcaaaa	taggctgtcc	4920
ccagtgcaag	tgcaggtgcc	agaacatttc	tctatcgaag	gatctgcgat	cgctccggtg	4980
cccgtcagtg	ggcagagcgc	acatcgccca	cagtcgccga	gaagttgggg	ggaggggtcg	5040
gcaattgaac	gggtgcctag	agaaggtggc	gcgggggtaaa	ctgggaaagt	gatgtcgtgt	5100
actggctccg	cctttttccc	gagggtgggg	gagaaccgta	tataagtga	gtagtgcgcg	5160
tgaacgttct	tttttcgcaac	gggtttgccc	ccagaacaca	gctgaagctt	cgaggggctc	5220
gcactctctc	ttcacgcgcc	cgccgcccta	cctgaggccg	ccatccacgc	cggttgagtc	5280
gcgtttctgcc	gcctcccgcg	tgtgtgtcct	cctgaactgc	gtccgcgcgt	taggtaagtt	5340
taaagctcag	gtcgagaccg	ggcctttgtc	cggcgcctcc	ttggagccta	cctagactca	5400
gccggctctc	cacgctttgc	ctgaccctgc	ttgttcaact	ctacgtcttt	gtttcgtttt	5460
ctgtttctgcg	ccgttacaga	tccaagctgt	gaccggcgcc	tacgtaagtg	atatctacta	5520
gatttatcaa	aaagagtgtt	gacttgtgag	cgctcacaat	tgatacttag	attcatcgag	5580
agggacacgt	cgactactaa	ccttcttctc	tttctctacag	ctgagat		5627

<210> 28
 <211> 553
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pMG plasmid from InvivoGen: EMCV IRES sequence

<400> 28						60
aacgttactg	gccgaagccg	cttggaaataa	ggccgggtgtg	cgtttgtota	tatgttattt	120
tccaccatat	tgccgtcttt	tggcaatgtg	agggcccgga	aacctggccc	tgtcttcttg	180
acgagcattc	ctaggggtct	ttcccctctc	gccaaaggaa	tgcaaggtct	gttgaatgtc	240
gtgaaggaag	cagttcctct	ggaagcttct	tgaagacaaa	caacgtctgt	agcgaccctt	300
tgcaggcagc	ggaaccccc	acctggcgac	aggtgcctct	gcggccaaaa	gccacgtgta	360
taagatacac	ctgcaaaggc	ggcacaaccc	cagtgcacag	ttgtgagttg	gatagttgtg	420
gaaagagtca	aatggctctc	ctcaagcgta	ttcaacaagg	ggctgaagga	tgcccagaag	480
gtaccccat	gtatgggatc	tgatctgggg	cctcgggtgca	catgctttac	gtgtgttttag	540
tcgagggttaa	aaaacgtcta	ggccccccga	accacggggga	cgtgggttttc	ctttgaaaaa	553
cacgatgata	ata					

<210> 29
 <211> 4692
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pDSred1-N1 plasmid from Clontech

<400> 29	tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg	60
----------	--	----

cgttacataa	cttacggttaa	atggcccgcg	tggttgaccg	cccaacgacc	cccgcccaatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacggt	aaactgcccc	cttggcgagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cgggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgt	acaactccgc	cccattgacg	caaattggcg	gtaggcgtgt	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacggtagc	gcgggcccg	660
gatccaccgg	tcgccaccat	ggtgcgctcc	tccaagacg	tcatacagga	gttcacgcgc	720
ttcaaggtgc	gcatggaggg	caccgtgaac	ggccacgagt	tcgagatcga	gggcgagggc	780
gagggccgccc	cctacgaggg	ccacaacacc	gtgaagctga	aggtgaccaa	gggcggcccc	840
ctgcccttcg	cctgggacat	cctgtccccc	cagttccagt	acggctccaa	ggtgtacgtg	900
aagcaccccc	ccgacatccc	cgactacaag	aagctgtcct	tccccgaggg	cttcaagtgg	960
gagcgcgtga	tgaacttcga	ggacggcggc	gtggtgaccg	tgaccacagga	ctcctccctg	1020
caggacggct	gcttcatcta	caaggtgaag	ttcatcgggc	tgaacttccc	ctccgacggc	1080
cccgtaatgc	agaagaagac	catgggctgg	gaggcctcca	ccgagcgctc	gtacccccgc	1140
gacggcgctgc	tgaaggcgga	gatccacaag	gcctctgtag	tgaaggacgg	gtgccactac	1200
ctggtggagt	tcaagtccat	ctacatggcc	aagaagcccc	tgacgctgcc	cggtacttac	1260
tacgtggact	ccaagctgga	catcacctcc	cacaacgagg	actacaccat	cgtggagcag	1320
tacgagcgca	ccgagggccg	ccaccacctg	ttcctgtagc	ggcgcgact	ctagatcata	1380
atcagccata	ccacatttgt	agaggtttta	cttgctttta	aaaacctccc	acacctcccc	1440
ctgaacctga	aacataaaaat	gaatgcaatt	gttggttgta	acttgtttat	tcgagcttat	1500
aatggttaca	ataaaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttcactg	1560
cattctagtt	gtggtttgtc	caaactcctc	aatgtatctt	aaggcgtaaa	ttgtaagcgt	1620
taatatTTTT	ttaaaattcg	cgttaaatTT	ttgttaaato	agctcatttt	ttaaccaata	1680
ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	accgagatag	ggttgagtgt	1740
tgctccagtt	tggaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaaaggcg	1800
aaaaaccgtc	tatcagggcg	atggcccact	acgtgaacca	tcaccctaata	caagtTTTT	1860
ggggtcgagg	tgccgtaaaag	cactaaatcg	gaaccctaaa	gggagcccc	gatttagagc	1920
ttgacggggg	aagccggcga	acgtggcgag	aaaggaaggg	aagaaagcga	aaggagccgg	1980
cgctagggcg	tagcgggtcac	gctgcgcgta	gctgcgcgta	accaccacac	ccgcccgcgt	2040
taatgcgcgc	ctacagggcg	cgctcaggtg	cacttttccg	ggaaatgtgc	gcggaacccc	2100
tatttgtttt	tttttctaaa	tacattcaaa	tatgtatccg	ctcatgagac	aataaccttg	2160
ataaatgctt	caataaatatt	gaaaaaggaa	gagtcctgag	gcggaagaaa	ccagctgtgg	2220
aatgtgtgtc	agttaggggtg	tggaaagtcc	ccaggctccc	cagcaggcag	aagtatgcaa	2280
agcatgcata	tcaatttagtc	agcaaccagg	tgtggaaagt	ccccaggctc	cccagcaggc	2340
agaagtatgc	aaagcatgca	tctcaattag	tgccaacca	tagtcccgc	cctaactccg	2400
cccatccgcg	ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	ctgactaatt	2460
ttttttatTT	atgcagaggg	cgaggccgccc	tcggcctctg	agctattcca	gaagtagtga	2520
ggaggctttt	ttggaggcct	aggcttttgc	aaagatcgat	caagagacag	gatgaggatc	2580
gtttcgcatg	attgaacaag	atggattgca	cgagggttct	ccggccgctt	gggtggagag	2640
gctattcggc	tatgactggg	cacaacagac	aatcggtgc	tctgatgccg	ccgtgttccg	2700
gctgtcagcg	cagggggcgcc	cgggttctttt	tgtcaagacc	gacctgtccg	gtgccctgaa	2760
tgaactgcaa	gacgaggcag	cgcggtctac	gtggctggcc	acgacggggc	ttccttgcgc	2820
agctgtgctc	gacgttgtca	ctgaagcggg	aaggggactg	ctgctattgg	gcgaagtgcc	2880
ggggcaggat	ctcctgtcat	ctcaccttgc	tcttgccgag	aaagtatcca	tcatggctga	2940
tgcaatgcgg	cggctgcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	3000
acatcgcatc	gagcagcac	gtactcggat	ggaagccggg	cttgctgatc	aggatgatct	3060
ggacgaagag	catcaggggc	tcgcgccagc	cgaactgttc	gccaggctca	aggcgagcat	3120
gcccgcaggg	gaggatctcg	tcgtgaccca	tggcgatgcc	tgcttgccga	atatcatggt	3180
ggaaaatggc	cgtttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggacgcgta	3240
tcaggacata	gcgttgggta	cccgtgatat	tgctgaagag	cttggcgggc	aatgggctga	3300
ccgcttccct	gtgctttacg	gtatcgccgc	tcccgattcg	cagcgcacgc	ccttctatcg	3360
ccttcttgac	gagttcttct	gagcgggact	ctgggggttcg	aaatgaccga	ccaagcgacg	3420
cccaacctgc	catcacgaga	tttcgattcc	accgcgcgct	tctatgaaag	gttgggcttc	3480
ggaatcgttt	tccgggacgc	cggctggatg	atcctccagc	gcggggatct	catgctggag	3540
ttcttcgccc	accctagggg	gaggctaact	gaaacacgga	aggagacaat	accggaagga	3600
accgcgccta	tgacggcaat	aaaaagacag	aataaaaacg	acggtgttgg	gtcgtttgtt	3660
cataaaacgc	gggttcggtc	ccagggtcgg	cactctgtcg	atacccacc	gagaccccat	3720
tggggccaat	acgcccgcgt	ttcttccctt	tccccacccc	accccccaag	ttcgggtgaa	3780
ggcccagggc	tcgcagccaa	cgtcggggcg	gcaggccctg	ccatagcctc	aggttactca	3840
tatatcattt	agatttgatt	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	3900
ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	3960
gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	cgtaatctgc	4020
tgcttgcaaa	caaaaaaac	accgctacca	gcggtggttt	gtttgcggga	tcaagagcta	4080

ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	4140
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	4200
gctctgctaa	tcctgttacc	agtggctgct	gccagtgggc	ataagtctgt	tcttaccggg	4260
ttggactcaa	gacgatagtt	accggataag	gcgagcgggt	cgggctgaac	gggggggttc	4320
tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	4380
ctatgagaaa	gcgccacgct	tcccgaaggg	agaaaggcgg	acaggtatcc	ggtaagcggc	4440
agggtcggaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	4500
agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	4560
gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	ggccttttgc	4620
tggccttttg	ctcacatggt	ctttcctgcg	ttatcccctg	attctgtgga	taaccgtatt	4680
accgccatgc	at					4692

<210> 30

<211> 4257

<212> DNA

<213> Artificial Sequence

<220>

<223> pPur plasmid from Clontech

<400> 30

ctgtggaatg	tgtgtcagtt	agggtgtgga	aagtccccag	gctccccagc	aggcagaagt	60
atgcaaagca	tgcattctca	ttagtccagc	accaggtgtg	gaaagtcccc	aggctcccc	120
gcaggcagaa	gtatgcaaag	catgcatctc	aattagtcag	caaccatagt	cccgccccta	180
actccgcccc	tcccgcacct	aactccgccc	agttccgccc	attctccgcc	ccatggctga	240
ctaatttttt	ttatttatgc	agaggccgag	gcgcctcgg	cctctgagct	attccagaag	300
tagtgaggag	gcttttttgg	aggcctaggg	ttttgcaaaa	agcttgcatg	cctgcaggtc	360
ggccgccacg	accggtgccc	ccaccaatccc	ctgaccacac	cccctgaccc	ctcacaagga	420
gacgaccttc	catgaccgag	tacaagcccc	cggtgcgcct	cgccacccgc	gacgacgtcc	480
cccgggcccgt	acgcaccttc	gccgcgcgct	tcgccgacta	ccccgccacg	cgccacaccc	540
tcgacccgga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	600
tcggggtcga	catcggcaag	gtgtgggtcg	cggacgacgg	cgccgcgggt	gcggtctgga	660
ccacgccgga	gagcgtcgaa	gcgggggccc	tgttcgccga	gacgggccc	cgcatggccc	720
agttgagcgg	ttcccggctg	gccgcgcagc	aacagatgga	aggcctcctg	gcgcccgaac	780
ggcccaagga	gcccgcgtgg	ttcctggcca	ccgtcggcgt	ctcgcccgaac	caccagggca	840
agggtctggg	cagcgcgcgt	gtgtcccccg	gagtgaggag	ggccgagcgc	gcccgggtgc	900
ccgccttcct	cgcccttcct	gcgcccgcga	acctcccctt	ctacgagcgg	ctcggcttca	960
ccgtcaccgc	cgacgtcgag	gtgcccgaag	gaccgcgcac	ctgggtgcatg	acccgcaagc	1020
ccggtgcctg	acgcccgcgc	cacgacccgc	agcgcgcgac	cgaaaggagc	gcacgacccc	1080
atggtcccca	ccgaagccga	cccgggcggc	cccgcgcgac	ccgcacccgc	ccccgaggcc	1140
caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcctta	1200
aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tgttgttgtt	1260
aacttgttta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	1320
aataaagcat	ttttttcact	gcattctagt	tgtgttttgt	ccaaactcat	caatgtatct	1380
tatcatgtct	ggatccccag	gaagctcctc	tgtgtcctca	taaaccttaa	cctcctctac	1440
ttgagaggac	attccaatca	taggctgccc	atccaccctc	tgtgtcctcc	tgttaattag	1500
gtcacttaac	aaaaaggaaa	ttgggtaggg	gtttttcaca	gaccgctttc	taagggtaat	1560
tttaaaatat	ctgggaagtc	ccttcacact	ctgtgttcca	gaagtgttgg	taaacagccc	1620
acaaatgtca	acagcagaaa	catacaagct	gtcagctttg	cacaaggggc	caacaccctg	1680
ctcatcaaga	agcactgtgg	ttgtctgtgt	agtaatgtgc	aaaacaggag	gcacattttc	1740
cccacctgtg	taggtttcaa	aatatctagt	gttttcattt	ttacttggat	caggaaccca	1800
gcactccact	ggataagcat	tatccttatc	caaaacagcc	ttgtgggtcag	tgttcatctg	1860
ctgactgtca	actgtagcat	tttttggggg	tacagtttga	gcaggatatt	tggtcctgta	1920
gtttgctaac	acacctgca	gctccaaaag	ttccccacca	acagcaaaaa	aatgaaaatt	1980
tgaccttgta	atgggttttc	cagcaccatt	ttcatgagtt	ttttgtgtcc	ctgaatgcaa	2040
gtttaacata	gcagttaccc	caataacctc	agttttaaca	gtaacagctt	cccacatcaa	2100
aatattttcca	cagggttaagt	cctcatttta	attaggcaaa	ggaattcttg	aagacgaaa	2160
ggcctcgtga	tacgcctatt	tttatagggt	aatgtcatga	taataatggg	ttcttagacg	2220
tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	tttctaaata	2280
cattcaataa	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	ataaatattga	2340
aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	ttttgoggca	2400
ttttgccttc	ctgtttttgc	tcacccagaa	acgctgtgta	aagtaaaaga	tgctgaagat	2460
cagttgggtg	cacgagtggt	ttacatcgaa	ctggatctca	acagcggtaa	gatccttgag	2520
agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	ttaaagtctc	gctatgtggc	2580
gcggtattat	cccggtgtga	cgccggggcaa	gagcaactcg	gtcgcgcgat	acactattct	2640
cagaatgact	tgggttagta	ctcaccagtc	acagaaaagc	atcttacgga	tggcatgaca	2700
gtaagagaat	tatgcagtcg	tgccataacc	atgagtata	acactgcggc	caacttactt	2760

ctgacaacga	tgggaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	2820
gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	2880
gacaccacga	tgctgcagc	aatggcaaca	acgttgcgca	aactattaac	tggcgaacta	2940
cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	agttgcagga	3000
ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	tggagccggg	3060
gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	3120
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	3180
gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	3240
cttttagattg	attttaaact	tcatTTTTaa	ttttaaagga	tctagggtgaa	gatccttttt	3300
gataatctca	tgaccaaact	cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	3360
gtagaaaaga	tcaaaggatc	ttcttgagat	cctttttttc	tgcgcgtaat	ctgctgcttg	3420
caaacaaaaa	aaccaccgct	accagcgggtg	gtttgtttgc	cggtatcaaga	gctaccaact	3480
ctttttccga	aggtaactgg	cttcagcaga	gcgcagatac	caaataactgt	ccttctagt	3540
tagccgtagt	tagggccacca	cttcaagaac	tctgtagcac	cgccacata	cctcgctctg	3600
ctaactcctgt	taccagtggc	tgctgccagt	ggcgataagt	cgtgtcttac	cggtgtggac	3660
tcaagacgat	agttaccgga	taaggcgcag	cggtcgggct	gaacgggggg	ttcgtgcaca	3720
cagccagct	tggagcgaac	gacctacacc	gaactgagat	acctacagcg	tgagctatga	3780
gaaagcgcca	cgcttcccga	aggagaaaag	gcggacaggt	atccggtaag	cggcaggggtc	3840
ggaacaggag	agcgcacgag	ggagcttcca	gggggaaacg	cctgggtatct	ttatagtcct	3900
gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	agggggggcgg	3960
agcctatgga	aaaacgccag	caacgcggcc	tttttacggt	tcttgccctt	ttgctggcct	4020
tttgctcaca	tggtctttcc	tcggttatcc	cctgattctg	tggataaccg	tattaccgcc	4080
tttagtgag	ctgataccgc	tcgcccgcagc	cgaacgacgc	agcgcagcga	gtcagtgagc	4140
gaggaaagcgg	aagagcgct	gatgcggtat	tttctcctta	cgcactctgtg	cggtatttca	4200
caccgcatat	ggtgcactct	cagtacaact	tgctctgatg	cgcgatagtt	aagccag	4257

<210> 31
 <211> 8136
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pWE15 cosmid vector

<300>
 <308> GenBank X65279
 <309> 1995-04-14

<400> 31						
ctatagtgag	tcggtattatg	cgcccgcgaa	ttcttgaaga	cgaaagggcc	tcgtgatacg	60
cctattttta	tagggttaatg	tcattgataat	aatggtttct	tagacgtcag	gtggcacttt	120
tcggggaaat	gtgcgcggaa	cccctatttg	tttatttttc	taaatacatt	caaatatgta	180
tccgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	240
gagtatccaa	catttaactg	tcgcccttat	tccctttttt	gcggcatttt	gcttctctgt	300
tttgctcacc	cagaaacgct	ggtgaaagta	aaagatgctg	aagatcagtt	gggtgcacga	360
gtgggttaca	tcgaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tcgccccgaa	420
gaacgttttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	480
gttgacgcgg	ggcaagagca	actcggctcg	cgcatacact	attctcagaa	tgacttgggt	540
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	600
agtgtctcca	taaccatgag	tgataaacct	gcggccaact	tacttctgac	aacgatcgga	660
ggaccggaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tcgccttgat	720
cgttgggaac	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	780
gcagcaatgg	caacaacggt	gcgcaaaact	ttacttggcg	aactacttac	tctagcttcc	840
cggcaacaat	taatagactg	gatggaggcg	gataaagtgt	caggaccact	tctgcgctcg	900
gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	960
ggtatcattg	cagcactggg	gccagatggg	aagccctccc	gtatcgtagt	tatctacacg	1020
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	1080
ctgatttaagc	attggttaact	gtcagaccaa	gtttactcat	atatacttta	gattgattta	1140
aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	1200
aaaatccctt	aacgtgagtt	ttcgtttcac	tgagcgtcag	accccgtaga	aaagatcaaa	1260
ggatcttctt	gagatccttt	tttctgcgc	gtaactctgt	gcttgcaaac	aaaaaaacca	1320
ccgctaccag	cgggtggtttg	tttgccggc	caagagctac	caactctttt	tccgaaggta	1380
actggcttca	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	gtagtttaggc	1440
caccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgctaata	cctgttacc	1500
gtggctgctg	ccagttggcg	taagtctgtg	cttaccgggt	tggactcaag	acgatagtt	1560
ccggataaag	cgcagcggtc	gggtgaaacg	gggggttcgt	gcacacagcc	cagcttggag	1620
cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaag	cggccagcct	1680

ccgaagggag	aaaggcggac	aggatccgg	taagcggcag	ggtcggaaca	ggagagcgca	1740
cgagggagct	tccaggggga	aacgcctggt	atctttatag	tcctgtcggg	gtttcgccac	1800
ctctgacttg	agcgtcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	1860
gccagcaacg	cggccttttt	acggttcctg	gccttttgc	ggccttttgc	tcacatgttc	1920
ttcctgcgt	tatccctga	ttctgtggat	aacggtatta	ccgcctttga	gtgagctgat	1980
accgctcgcc	gcagccgaac	gaccgagcgc	agcgagtcag	tgagcgagga	agcgggaagag	2040
cgctgacttc	cgcgtttcca	gactttacga	aacacggaaa	ccgaagacca	ttcatgttgt	2100
tgctcaggtc	gcagacgttt	tgcagcagca	gtcgttcac	gttcgctcgc	gtatcggtga	2160
ttcattctgc	taaccagtaa	ggcaaccccg	ccagcctagc	cgggtcctca	acgacaggag	2220
cacgatcatg	cgcacccgtc	agatccagac	atgataagat	acattgatga	gtttggacaa	2280
accacaacta	gaatgcagtg	aaaaaaatgc	tttatttgtg	aaatttgtga	tgctattgct	2340
ttatttghta	ccattataag	ctgcaataaa	caagttaaca	acaacaattg	cattcatttt	2400
atgtttcagg	ttcaggggga	gggtgtgggag	gttttttaaa	gcaagtaaaa	cctctacaaa	2460
tgtggtatgg	ctgattatga	tctctagtca	aggcaactata	catcaaata	tccttattaa	2520
ccccctttaca	aattaaaaag	ctaaagggtac	acaatttttg	agcatagtta	ttaatagcag	2580
acactctatg	cctgtgtgga	gtaagaaaaa	acagtatgtt	atgattataa	ctgttatgcc	2640
tacttataaa	ggttacagaa	tatttttcca	taattttctt	gtatagcagt	gcagcttttt	2700
cctttgtggt	gtaaatagca	aagcaagcaa	gagttctatt	actaaacaca	gcagtactca	2760
aaaaacttag	caattctgaa	ggaaagtcct	tgggtcttcc	tacctttctc	ttcttttttg	2820
gaggagttag	atgttgagag	tcagcagtag	cctcatcatc	actagatggc	atttcttctg	2880
agcaaaacag	gttttcctca	ttaaaggcat	tcaccaactg	ctccatttca	tcagttccat	2940
agggttgaat	ctaaaatata	caaacaatta	gaatcagtag	tttaacacat	tatacactta	3000
aaaattttat	atttacctta	gagctttaaa	tctctgtagg	tagtttgtcc	aattatgtca	3060
caccacagaa	gtaagggttcc	ttcacaaga	tcgggaccac	agcggccatc	gtgcctcccc	3120
actcctgcag	ttcgggggca	tggatgcgcg	gatagccgct	gctgggttcc	tggatgccga	3180
cggatttgca	ctgcgcgtag	aactcgcgag	gtcgtccagc	ctcaggcagc	agctgaacca	3240
actcgcgagg	ggatcgagcc	cggggtgggc	gaagaactcc	agcatgagat	ccccgcgctg	3300
gaggatcatc	cagccggcgt	cccggaaaac	gattccgaag	cccaaccttt	catagaaggc	3360
ggcggtggaa	tcgaaatctc	gtgatggcag	gttgggcgtc	gcttgggtcgg	tcatttcgaa	3420
ccccagatgc	ccgctcagaa	gaactcgtca	agaaggcgat	agaaggcgat	gcgctgcgaa	3480
tcgggagcgg	cgataccgta	aagcacgagg	aagcggtcag	cccattcgcc	gccaaagctc	3540
tcagcaatat	cacgggtagc	caacgctatg	tcctgatagc	ggtcogccac	acccagccgg	3600
ccacagtcga	tgaatccaga	aaagcggcca	ttttccacca	tgatattcgg	caagcaggca	3660
tcgccatggg	tcacgacgag	atcctcgccg	tcgggatgcg	cgccttgagc	ctggcgaaac	3720
gttcggctgg	cgcgagcccc	tgatgctctt	cgtccagatc	atcctgatcg	acaagaccgg	3780
cttccatccg	agtacgtgct	cgtcgtatgc	gatgtttcgc	ttggtgggtcg	aatgggcagg	3840
tagccggatc	aagcgtatgc	agccgcccga	ttgcatcagc	catgatggat	actttctcgg	3900
caggagcaag	gtgagatgac	aggagatcct	gccccggcac	ttcgcccaat	agcagccagt	3960
cccttcccgc	ttcagtgaca	acgtcgagca	cagctgcgca	aggaacgcc	gtcgtggcca	4020
ggcacgagga	ccgcgctgcc	tcgtcctgca	gttcattcag	ggcacccggac	aggtcggctc	4080
tgacaaaaag	aaccggggcg	ccctgcgctg	acagccggaa	cacggcggca	tcagagcagc	4140
cgattgtctg	ttgtgcccag	tcatagcgca	atagcctctc	cacccaagcg	gccgggagaa	4200
ctgcgtgcaa	tcacatctgt	tcaatcatgc	gaaacgatcc	tcacatctgtc	tcttgatcag	4260
atcttgatcc	cctgcgccat	cagatccttg	gcggcaagaa	agccatccag	tttactttgc	4320
agggcttccc	aaccttacca	gagggcgccc	cagctggcaa	ttccgggttcg	cttgctgtcc	4380
ataaaaccgc	ccagcttagc	tatcgccatg	taagcccact	gcaagctacc	tgctttctct	4440
ttgcgcttgc	gttttccott	gtccagatag	cccagtagct	gacattccatc	cgggggtcagc	4500
accgtttctg	cggactggct	ttctacgtgt	tcogcttcc	ttagcagccc	ttgcgcccctg	4560
agtgccttgc	gcagcgtgaa	agctttttgc	aaaagcctag	gcctccaaaa	aagcctcctc	4620
actacttctg	gaatagctca	gagggcggag	cggcctaaat	aaaaaaaatt	agtcagccat	4680
ggggcgagga	atggggcgga	ctggggcgag	ttagggggcg	gatggggcgga	gttagggggcg	4740
ggactatggt	tgctgactaa	ttgagatgca	tgctttgcat	acttctgcct	gctggggagc	4800
ctggggactt	tcacacacct	gttgctgact	aattgagatg	catgctttgc	atacttctgc	4860
ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	agccggatct	4920
gcaggaccga	acgctgccc	agatgcgcc	cgtgcggctg	ctggagatgg	cggacgcgat	4980
ggatattgtc	tgccaagggt	tggtttgcgc	attcacagtt	ctccgcaaga	attgattggc	5040
tcacattctt	ggagtggtag	atccgttagc	gaggtgcgc	cggcttccat	tcaggtcgag	5100
gtggccgggc	tcctatgcac	gcgacgcaac	gcggggaggc	agacaaggta	tagggcgggcg	5160
cctacaatcc	atgccaaccc	gttccatgtg	ctcgccgagg	cgcataaatc	gccgtgacga	5220
tcagcggtcc	aatgatcgaa	gttaggctgg	taagagccgc	gagcgatcct	tgaagctgtc	5280
cctgatggtc	tgcatctacc	tgccctggaca	gcattggcctg	caacgcggca	tcccgatgcc	5340
gccggaagcg	agaagaatca	taatggggaa	ggccatccag	cctcgctcgc	cgaacgccag	5400
caagacgtag	ccagcgcggt	cgggccggcca	tgccggcgat	aatggcctgc	ttctcgccga	5460
aacgtttggt	ggcgggacca	gtgacgaagg	cttgagcgag	ggcgtgcaag	attccgaata	5520
ccgcaagcga	caggccgatc	atcgtcgccg	tcagcgaaaa	gcggtcctcg	ccgaaaaatga	5580
ccagagcgcc	tgccggcacc	tgctctacga	gttgcatgat	aaagaagaca	gtcataagtg	5640
cggcgacgat	agtcatgccc	cgcgccacc	ggaaggagct	gactgggttg	aaggctctca	5700

agggcatcgg	tgcacgctct	cccttatgcg	actcctgcat	taggaagcag	cccagtagta	5760
gggttgaggcc	gttgagcacc	gccgcccgc	ggaatgggtgc	atgcaaggag	atggcgccca	5820
acagtgcccc	ggccacgggc	ctgccaccat	acccacgccc	aaacaagcgc	tcatgagccc	5880
gaagtggcga	gcccgatctt	ccccatcggg	gatgtcggcg	atataggcgc	cagcaaccgc	5940
acctgtggcg	cgggtgatgc	cggccacgat	gcgtccggcg	tagaggatct	tggcagtcac	6000
agcatgcgca	tatccatgct	tgcaccatgc	gctcacaag	taggtgaatg	cgcaatgtag	6060
taccacatc	gtcatcgctt	tccactgctc	tcgcgaataa	agatggaaaa	tcaatctcat	6120
ggtaatagtc	catgaaaatc	cttgtattca	taaatcctcc	aggtagctat	atgcaaattg	6180
aaacaaaaga	gatgggtgatc	tttctaagag	atgatggaat	ctcccttcag	tatcccgatg	6240
gtcaatgcgc	tggatatggg	atagatggga	atatgctgat	ttttatggga	cagagttgcg	6300
aactgttccc	aactaaaatc	attttgcacg	atcagcgcac	tacgaacttt	accacaaaat	6360
agtcaggtaa	tgaatcctga	tataaagaca	ggttgataaa	tcagtcttct	acgcgcacgc	6420
caacgcgcaca	cgcgtgaaag	tctttcagtt	gtgagcctgg	gcaaaccgtt	aactttcggc	6480
ggctttgctg	tgcgacaggc	tcacgtctaa	aaggaaataa	atcatgggtc	ataaaaattat	6540
cacgttgtcc	ggcgccggcg	cggatgttct	gtatgcgctg	tttttccgtg	gcgcgttgct	6600
gtctgggtgat	tttgcctcta	aatctggcac	agccgaattg	cgcgagcttg	gttttgctga	6660
aaccagacac	acagcaactg	aataccagaa	agaaaatcac	tttaccttct	tgacatcaga	6720
agggcagaaa	tttgccgttg	aacacctggg	caatcgcgct	tttgggtgagc	agcaatattg	6780
cgcttcgatg	acgcttggcg	ttgagattga	tacctctgct	gcacaaaagg	caatcgacga	6840
gctggaccag	cgcattctgc	acacgcgtct	cttcgaactt	attcgcaatg	gagtgctatt	6900
catcaaggac	gccgctatcg	caaatgggtgc	tatccacgca	gcggcaatcg	aaacacctca	6960
gccggtgacc	aataatctaca	acatcagcct	tggtatccag	cgtgatgagc	cagcgagaaa	7020
caaggtaaacc	gtcagtgccg	ataagttcaa	agttaaacct	ggtggttgata	ccaacattga	7080
aacgttgatc	gaaaacgcgc	tgaaaaacgc	tgctgaatgt	gcggcgctgg	atgtcacaaa	7140
gcaaatggca	gcagacaaga	aagcgatgga	tgaactggct	tcctatgtcc	gcacggccat	7200
catgatggaa	tgtttccccg	gtgggtgttat	ctggcagcag	tgccgtcgat	agtatgcaat	7260
tgataattat	tatcatttgc	gggtcctttc	cgcgcatccg	ccttgttacg	ggcgggcgac	7320
ctcgcgggtt	ttcgctattt	atgaaaattt	tcgggtttaa	ggcggttccg	ttcttcttcg	7380
tcataactta	atgtttttat	ttaaaatacc	ctctgaaaag	aaaggaaaacg	acaggtgctg	7440
aaagcgagct	ttttggcctc	tgctggtttcc	tttctctggt	tttgtccgtg	gaatgaacaa	7500
tggaggtcaa	caaaaagcag	ctggctgaca	ttttcgggtg	gagtatccgt	accattcaga	7560
actggcagga	acagggaatg	cccgttctgc	gaggcggtgg	caagggtaat	gaggtgcttt	7620
atgactctgc	cgccgtcata	aaatgggtatg	ccgaaaaggga	tgctgaaatt	gagaacgaaa	7680
agctgcgccg	ggaggttgaa	gaactgcggc	aggccagcga	ggcagatcca	caggacgggt	7740
gtggtcgcca	tgatcgcgta	gtcgatagtg	gctccaagta	gcgaagcgag	caggactggg	7800
cggcggaaca	gcggtcggac	agtgtccga	gaacgggtgc	gcatagaaat	tgcatcaacg	7860
catatagcgc	tagcagcacg	ccatagtgac	tgccgatgct	gtcggaaatg	acgatatccc	7920
gcaagaggcc	cggcagtagc	ggcataacca	agcctatgcc	tacagcatcc	aggggtgacgg	7980
tgccgaggat	gacgatgagc	gcattgttag	atttcataca	cgggtgcctga	ctgcgttagc	8040
aatttaactg	tgataaacta	cgcattaaa	gcttatcgat	gataagcggg	caaacatgag	8100
aattcgccgc	cgcaattaac	cctcactaaa	ggatcc			8136

<210> 32
 <211> 2713
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pNEB193 plasmid

<400> 32						
tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccc	tcaggggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cgccatcaga	gcagattgta	ctgagagtgc	180
accatattgc	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataaccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aaactgttgg	aaggggcgatc	ggtgcggggcc	tcttcgctat	300
tacgccagct	ggcgaaaagg	ggatgtgctg	caaggcgatt	aagtgggta	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgaatt	cgagctcggt	acccgggggc	420
gcgcgggatc	cttaattaag	tctagagtgc	actgtttaaa	cctgcaggca	tgcaagcttg	480
gcgtaatcat	ggatcatgct	gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	540
aacatacgag	ccggaagcat	aaagtgtaaa	gcctgggggtg	cctaagtgtg	gagctaactc	600
acattaattg	cgttgcgtc	actgcccgct	ttccagtcgg	gaaacctgtc	gtgccagctg	660
cattaatgaa	tcggccaacg	cgcggggaga	ggcgggtttgc	gtattgggcg	ctcttccgct	720
tctcgcgtca	ctgactcgct	gcgctcggtc	gttcgggtgc	ggcgagcggg	atcagctcac	780
tcaaaaggcg	taatacgggt	atccacagaa	tcagggggata	acgcaggaaa	gaacatgtga	840
gcaaaaggcc	agcaaaaggc	caggaaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	900
aggctccgcc	cccctgacga	gcatcacaaa	aatcgacgct	caagtacagag	gtggcgaaaac	960

ccgacaggac	tataaagata	ccaggcggtt	ccccctggaa	gctccctcgt	gcgctctcct	1020
gttccgaccc	tgccgcttac	cggataacctg	tccgcctttc	tcccttcggg	aagcgtggcg	1080
ctttctcata	gctcacgctg	taggtatctc	agttcgggtg	aggtcgttcg	ctccaagctg	1140
ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	1200
cttgagtcca	acccggttaag	acacgactta	tcgccactgg	cagcagccac	tggtaacagg	1260
attagcagag	cgaggtatgt	aggcggtgct	acagagttct	tgaagtgggtg	gcctaactac	1320
ggctacacta	gaaggacagt	atttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	1380
aaaagagttg	gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggttttttt	1440
gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgcctttt	1500
tctacggggt	ctgacgctca	gtggaacgaa	aactcacggt	aagggatttt	ggtcatgaga	1560
ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	1620
taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	gcttaatcag	tgaggcacct	1680
atctcagcga	tctgtctatt	tcgttcatcc	atagttgcct	gactccccgt	cgtgtagata	1740
actacgatac	gggagggctt	accatctggc	cccagtgctg	caatgatacc	gcgagaccca	1800
cgctcacccg	ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	1860
agtggttctg	caactttatc	cgctccatc	cagtctatta	attggtgccc	ggaagctaga	1920
gtaagtagtt	cgccagttaa	tagtttgccg	aacggtgttg	ccattgctac	aggcatcggt	1980
gtgtcacgct	cgctggttgg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	2040
gttacatgat	ccccatggt	gtgcaaaaaa	gcggttagct	ccttcggtcc	tccgatcggt	2100
gtcagaagta	agttggccgc	agtgttatca	ctcatgggta	tggcagcact	gcataattct	2160
cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	2220
ttctgagaat	agtgtatgcg	gcgaccgagt	tgcctctgcc	cggcgctcaat	acgggataat	2280
accgcgccac	atagcagaac	tttaaaagtg	ctcatcattg	gaaaacgttc	ttcggggcga	2340
aaactctcaa	ggatcttacc	gctgttgaga	tccagttcga	tgtaaccac	tcgtgcaccc	2400
aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	2460
caaaatgccg	caaaaaaggg	aataagggcg	acacgggaaat	ggtgaatact	catactcttc	2520
ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatattt	2580
gaatgtatct	agaaaaataa	acaaataggg	gttccgcgca	catttccccg	aaaagtgcc	2640
cctgacgtct	aagaaacctat	tattatcatg	acattaacct	ataaaaaatag	gcgtatcacg	2700
aggccctttc	gtc					2713

<210> 33
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attP

<400> 33
 cagctttttt atactaagtt g 21

<210> 34
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attB

<400> 34
 ctgctttttt atactaactt g 21

<210> 35
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> attL

<400> 35
 ctgctttttt atactaagtt g 21

<210> 36
 <211> 21
 <212> DNA

<213> Artificial Sequence

<220>

<223> attR

<400> 36

cagctttttt atactaactt g

21

<210> 37

<211> 1071

<212> DNA

<213> Artificial Sequence

<220>

<223> Integrase E174R

<221> CDS

<222> (1)...(1071)

<223> Nucleotide sequence encoding Integrase E147R

<400> 37

atg gga aga agg cga agt cat gag cgc cgg gat tta ccc cct aac ctt	48
Met Gly Arg Arg Ser His Glu Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
tat ata aga aac aat gga tat tac tgc tac agg gac cca agg acg ggt	96
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
aaa gag ttt gga tta ggc aga gac agg cga atc gca atc act gaa gct	144
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
ata cag gcc aac att gag tta ttt tca gga cac aaa cac aag cct ctg	192
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
aca gcg aga atc aac agt gat aat tcc gtt acg tta cat tca tgg ctt	240
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
gat cgc tac gaa aaa atc ctg gcc agc aga gga atc aag cag aag aca	288
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	
85 90 95	
ctc ata aat tac atg agc aaa att aaa gca ata agg agg ggt ctg cct	336
Leu Ile Asn Tyr Met Ser Lys Ile Lys Ala Ile Arg Arg Gly Leu Pro	
100 105 110	
gat gct cca ctt gaa gac atc acc aca aaa gaa att gcg gca atg ctc	384
Asp Ala Pro Leu Glu Asp Ile Thr Thr Lys Glu Ile Ala Ala Met Leu	
115 120 125	
aat gga tac ata gac gag ggc aag gcg gcg tca gcc aag tta atc aga	432
Asn Gly Tyr Ile Asp Glu Gly Lys Ala Ala Ser Ala Lys Leu Ile Arg	
130 135 140	
tca aca ctg agc gat gca ttc cga gag gca ata gct gaa ggc cat ata	480
Ser Thr Leu Ser Asp Ala Phe Arg Glu Ala Ile Ala Glu Gly His Ile	
145 150 155 160	
aca aca aac cat gtc gct gcc act cgc gca gca aaa tct aga gta agg	528
Thr Thr Asn His Val Ala Ala Thr Arg Ala Ala Lys Ser Arg Val Arg	
165 170 175	
aga tca aga ctt acg gct gac gaa tac ctg aaa att tat caa gca gca	576
Arg Ser Arg Leu Thr Ala Asp Glu Tyr Leu Lys Ile Tyr Gln Ala Ala	

180										185										190										
gaa	tca	tca	cca	tgt	tgg	ctc	aga	ctt	gca	atg	gaa	ctg	gct	gtt	gtt		624													
Glu	Ser	Ser	Pro	Cys	Trp	Leu	Arg	Leu	Ala	Met	Glu	Leu	Ala	Val	Val															
		195					200					205																		
acc	ggg	caa	cga	gtt	ggg	gat	tta	tgc	gaa	atg	aag	tgg	tct	gat	atc		672													
Thr	Gly	Gln	Arg	Val	Gly	Asp	Leu	Cys	Glu	Met	Lys	Trp	Ser	Asp	Ile															
	210					215					220																			
gta	gat	gga	tat	ctt	tat	gtc	gag	caa	agc	aaa	aca	ggc	gta	aaa	att		720													
Val	Asp	Gly	Tyr	Leu	Tyr	Val	Glu	Gln	Ser	Lys	Thr	Gly	Val	Lys	Ile															
	225				230					235					240															
gcc	atc	cca	aca	gca	ttg	cat	att	gat	gct	ctc	gga	ata	tca	atg	aag		768													
Ala	Ile	Pro	Thr	Ala	Leu	His	Ile	Asp	Ala	Leu	Gly	Ile	Ser	Met	Lys															
				245					250					255																
gaa	aca	ctt	gat	aaa	tgc	aaa	gag	att	ctt	ggc	gga	gaa	acc	ata	att		816													
Glu	Thr	Leu	Asp	Lys	Cys	Lys	Glu	Ile	Leu	Gly	Gly	Glu	Thr	Ile	Ile															
			260					265					270																	
gca	tct	act	cgt	cgc	gaa	ccg	ctt	tca	tcc	ggc	aca	gta	tca	agg	tat		864													
Ala	Ser	Thr	Arg	Arg	Glu	Pro	Leu	Ser	Ser	Gly	Thr	Val	Ser	Arg	Tyr															
		275					280					285																		
ttt	atg	cgc	gca	cga	aaa	gca	tca	ggg	ctt	tcc	ttc	gaa	ggg	gat	ccg		912													
Phe	Met	Arg	Ala	Arg	Lys	Ala	Ser	Gly	Leu	Ser	Phe	Glu	Gly	Asp	Pro															
	290					295					300																			
cct	acc	ttt	cac	gag	ttg	cgc	agt	ttg	tct	gca	aga	ctc	tat	gag	aag		960													
Pro	Thr	Phe	His	Glu	Leu	Arg	Ser	Leu	Ser	Ala	Arg	Leu	Tyr	Glu	Lys															
	305				310					315					320															
cag	ata	agc	gat	aag	ttt	gct	caa	cat	ctt	ctc	ggg	cat	aag	tgc	gac		1008													
Gln	Ile	Ser	Asp	Lys	Phe	Ala	Gln	His	Leu	Leu	Gly	His	Lys	Ser	Asp															
				325					330					335																
acc	atg	gca	tca	cag	tat	cgt	gat	gac	aga	ggc	agg	gag	tgg	gac	aaa		1056													
Thr	Met	Ala	Ser	Gln	Tyr	Arg	Asp	Asp	Arg	Gly	Arg	Glu	Trp	Asp	Lys															
			340					345					350																	
att	gaa	atc	aaa	taa													1071													
Ile	Glu	Ile	Lys	*																										
			355																											

<210> 38
 <211> 356
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Integrase E147R

<400> 38
 Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu
 1 5 10
 Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly
 20 25 30
 Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala
 35 40 45
 Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu
 50 55 60
 Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu
 65 70 75 80
 Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr

Leu	Ile	Asn	Tyr	85	Met	Ser	Lys	Ile	Lys	90	Ala	Ile	Arg	Arg	Gly	95	Leu	Pro
Asp	Ala	Pro	Leu	100	Glu	Asp	Ile	Thr	Thr	105	Lys	Glu	Ile	Ala	Ala	Met	Leu	
Asn	Gly	Tyr	Ile	115	Asp	Glu	Gly	Lys	Ala	120	Ala	Ser	Ala	Lys	Leu	Ile	Arg	
Ser	Thr	Leu	Ser	130	Asp	Ala	Phe	Arg	Glu	135	Ala	Ile	Ala	Glu	Gly	His	Ile	
Thr	Thr	Asn	His	145	Val	Ala	Ala	Thr	Arg	150	Ala	Ala	Lys	Ser	Arg	Val	Arg	
Arg	Ser	Arg	Leu	165	Thr	Ala	Asp	Glu	Tyr	170	Leu	Lys	Ile	Tyr	Gln	Ala	Ala	
Glu	Ser	Ser	Pro	180	Cys	Trp	Leu	Arg	Leu	185	Ala	Met	Glu	Leu	Ala	Val	Val	
Thr	Gly	Gln	Arg	195	Val	Gly	Asp	Leu	Cys	200	Glu	Met	Lys	Trp	Ser	Asp	Ile	
Val	Asp	Gly	Tyr	210	Leu	Tyr	Val	Glu	Gln	215	Ser	Lys	Thr	Gly	Val	Lys	Ile	
Ala	Ile	Pro	Thr	225	Ala	Leu	His	Ile	Asp	230	Ala	Leu	Gly	Ile	Ser	Met	Lys	
Glu	Thr	Leu	Asp	245	Lys	Cys	Lys	Glu	Ile	250	Leu	Gly	Gly	Glu	Thr	Ile	Ile	
Ala	Ser	Thr	Arg	260	Arg	Glu	Pro	Leu	Ser	265	Ser	Ser	Gly	Thr	Val	Ser	Arg	
Phe	Met	Arg	Ala	275	Arg	Lys	Ala	Ser	Gly	280	Leu	Ser	Phe	Glu	Gly	Asp	Pro	
Pro	Thr	Phe	His	290	Glu	Leu	Arg	Ser	Leu	295	Ser	Ala	Arg	Leu	Tyr	Glu	Lys	
Gln	Ile	Ser	Asp	305	Lys	Phe	Ala	Gln	His	310	Leu	Leu	Gly	His	Lys	Ser	Asp	
Thr	Met	Ala	Ser	325	Gln	Tyr	Arg	Asp	Asp	330	Arg	Gly	Arg	Glu	Trp	Asp	Lys	
Ile	Glu	Ile	Lys	340						345								
				355														

<210> 39
 <211> 876
 <212> DNA
 <213> Discosoma species

<220>
 <221> CDS
 <222> (45)...(737)
 <223> Nucleotide sequence encoding red flourescent protein (FP593)

<300>
 <308> GenBank AF272711
 <309> 2000-09-26

<400> 39
 agtttcagcc agtgacaggg tgagctgccca ggtattctaa caag atg agt tgt tcc 56
 Met Ser Cys Ser
 1

aag aat gtg atc aag gag ttc atg agg ttc aag gtt cgt atg gaa gga 104
 Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val Arg Met Glu Gly
 5 10 15 20

acg gtc aat ggg cac gag ttt gaa ata aaa ggc gaa ggt gaa ggg agg 152
 Thr Val Asn Gly His Glu Phe Glu Ile Lys Gly Glu Gly Glu Gly Arg
 25 30 35

cct tac gaa ggt cac tgt tcc gta aag ctt atg gta acc aag ggt gga 200
 Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val Thr Lys Gly Gly

40										45										50										
cct	ttg	cca	ttt	gct	ttt	gat	att	ttg	tca	cca	caa	ttt	cag	tat	gga	248														
Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln	Phe	Gln	Tyr	Gly															
		55					60					65																		
agc	aag	gta	tat	gtc	aaa	cac	cct	gcc	gac	ata	cca	gac	tat	aaa	aag	296														
Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro	Asp	Tyr	Lys	Lys															
	70					75					80																			
ctg	tca	ttt	cct	gag	gga	ttt	aaa	tgg	gaa	agg	gtc	atg	aac	ttt	gaa	344														
Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val	Met	Asn	Phe	Glu															
	85				90				95						100															
gac	ggg	ggc	gtg	gtt	act	gta	tcc	caa	gat	tcc	agt	ttg	aaa	gac	ggc	392														
Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser	Leu	Lys	Asp	Gly															
				105				110						115																
tgt	ttc	atc	tac	gag	gtc	aag	ttc	att	ggg	gtg	aac	ttt	cct	tct	gat	440														
Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn	Phe	Pro	Ser	Asp															
			120					125					130																	
gga	cct	gtt	atg	cag	agg	agg	aca	cgg	ggc	tgg	gaa	gcc	agc	tct	gag	488														
Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu	Ala	Ser	Ser	Glu															
		135					140					145																		
cgt	ttg	tat	cct	cgt	gat	ggg	gtg	ctg	aaa	gga	gac	atc	cat	atg	gct	536														
Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp	Ile	His	Met	Ala															
	150					155					160																			
ctg	agg	ctg	gaa	gga	ggc	ggc	cat	tac	ctc	gtt	gaa	ttc	aaa	agt	att	584														
Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu	Phe	Lys	Ser	Ile															
	165				170				175						180															
tac	atg	gta	aag	aag	cct	tca	gtg	cag	ttg	cca	ggc	tac	tat	tat	gtt	632														
Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly	Tyr	Tyr	Tyr	Val															
				185					190						195															
gac	tcc	aaa	ctg	gat	atg	acg	agc	cac	aac	gaa	gat	tac	aca	gtc	gtt	680														
Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp	Tyr	Thr	Val	Val															
			200					205					210																	
gag	cag	tat	gaa	aaa	acc	cag	gga	cgc	cac	cat	ccg	ttc	att	aag	cct	728														
Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro	Phe	Ile	Lys	Pro															
		215					220					225																		
ctg	cag	tga	actcggctca	gtcatggatt	agcggtaatg	gccacaaaag										777														
Leu	Gln	*																												
		230																												
gcacgatgat	cgtttttttag	gaatgcagcc	aaaaattgaa	ggttatgaca	gtagaaatac	837																								
aagcaacagg	ctttgcttat	taaacatgta	attgaaaac			876																								

<210> 40

<211> 230

<212> PRT

<213> Discosoma species

<400> 40

Met	Ser	Cys	Ser	Lys	Asn	Val	Ile	Lys	Glu	Phe	Met	Arg	Phe	Lys	Val
1				5					10					15	
Arg	Met	Glu	Gly	Thr	Val	Asn	Gly	His	Glu	Phe	Glu	Ile	Lys	Gly	Glu
			20					25					30		
Gly	Glu	Gly	Arg	Pro	Tyr	Glu	Gly	His	Cys	Ser	Val	Lys	Leu	Met	Val
		35				40						45			
Thr	Lys	Gly	Gly	Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln
	50					55					60				

```

Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His Pro Ala Asp Ile Pro
65          70          75          80
Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val
85          90          95
Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Ser Gln Asp Ser Ser
100        105        110
Leu Lys Asp Gly Cys Phe Ile Tyr Glu Val Lys Phe Ile Gly Val Asn
115        120        125
Phe Pro Ser Asp Gly Pro Val Met Gln Arg Arg Thr Arg Gly Trp Glu
130        135        140
Ala Ser Ser Glu Arg Leu Tyr Pro Arg Asp Gly Val Leu Lys Gly Asp
145        150        155        160
Ile His Met Ala Leu Arg Leu Glu Gly Gly Gly His Tyr Leu Val Glu
165        170        175
Phe Lys Ser Ile Tyr Met Val Lys Lys Pro Ser Val Gln Leu Pro Gly
180        185        190
Tyr Tyr Tyr Val Asp Ser Lys Leu Asp Met Thr Ser His Asn Glu Asp
195        200        205
Tyr Thr Val Val Glu Gln Tyr Glu Lys Thr Gln Gly Arg His His Pro
210        215        220
Phe Ile Lys Pro Leu Gln
225          230

```

```

<210> 41
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-att;

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

```

<400> 41
rkycwgcttt yktrtacnaa stsgb

```

25

```

<210> 42
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-attB;

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```

```

<400> 42
agccwgcttt yktrtacnaa ctsgb

```

25

```

<210> 43
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> m-attR

```

```

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

```


<400> 43
gttcagcttt cktrtacnaa ctsgb 25

<210> 44
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attL

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 44
agccwgcttt cktrtacnaa gtsgb 25

<210> 45
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> m-attP1

<221> misc_difference
<222> 18
<223> n is a or g or c or t/u

<400> 45
gttcagcttt yktrtacnaa gtsgb 25

<210> 46
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB1

<400> 46
agcctgcttt tttgtacaaa cttgt 25

<210> 47
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB2

<400> 47
agcctgcttt cttgtacaaa cttgt 25

<210> 48
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attB3

<400> 48
accagcttt cttgtacaaa cttgt 25

<210> 49

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR1

<400> 49
gttcagcttt tttgtacaaa cttgt 25

<210> 50
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR2

<400> 50
gttcagcttt cttgtacaaa cttgt 25

<210> 51
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attR3

<400> 51
gttcagcttt cttgtacaaa gttgg 25

<210> 52
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL1

<400> 52
agcctgcttt tttgtacaaa gttgg 25

<210> 53
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL2

<400> 53
agcctgcttt cttgtacaaa gttgg 25

<210> 54
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attL3

<400> 54
accagcttt cttgtacaaa gttgg 25

<210> 55
<211> 25

<212> DNA
<213> Artificial Sequence

<220>
<223> attP1

<400> 55
gttcagcttt tttgtacaaa gttgg 25

<210> 56
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> attP2,P3

<400> 56
gttcagcttt cttgtacaaa gttgg 25

<210> 57
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Lox P site

<400> 57
ataacttcgt ataatgtatg ctatacgaag ttat 34

<210> 58
<211> 1032
<212> DNA
<213> Escherichia coli

<220>
<221> CDS
<222> (1)...(1032)
<223> nucleotide sequence encoding Cre recombinase

<400> 58
atg tcc aat tta ctg acc gta cac caa aat ttg cct gca tta ccg gtc 48
Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1 5 10 15

gat gca acg agt gat gag gtt cgc aag aac ctg atg gac atg ttc agg 96
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
20 25 30

gat cgc cag gcg ttt tct gag cat acc tgg aaa atg ctt ctg tcc gtt 144
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
35 40 45

tgc cgg tcg tgg gcg gca tgg tgc aag ttg aat aac cgg aaa tgg ttt 192
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
50 55 60

ccc gca gaa cct gaa gat gtt cgc gat tat ctt cta tat ctt cag gcg 240
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
65 70 75 80

cgc ggt ctg gca gta aaa act atc cag caa cat ttg ggc cag cta aac 288
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
85 90 95

atg ctt cat cgt cgg tcc ggg ctg cca cga cca agt gac agc aat gct 336

Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	Asn	Ala		
			100					105						110			
gtt	tca	ctg	gtt	atg	cgg	cgg	atc	cga	aaa	gaa	aac	gtt	gat	gcc	ggg	384	
Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	Ala	Gly		
		115					120					125					
gaa	cgt	gca	aaa	cag	gct	cta	gcg	ttc	gaa	cgc	act	gat	ttc	gac	cag	432	
Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	Asp	Gln		
	130					135					140						
gtt	cgt	tca	ctc	atg	gaa	aat	agc	gat	cgc	tgc	cag	gat	ata	cgt	aat	480	
Val	Arg	Ser	Leu	Met	Glu	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	Arg	Asn		
	145				150					155					160		
ctg	gca	ttt	ctg	ggg	att	gct	tat	aac	acc	ctg	tta	cgt	ata	gcc	gaa	528	
Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	Ala	Glu		
			165					170						175			
att	gcc	agg	atc	agg	gtt	aaa	gat	atc	tca	cgt	act	gac	ggg	ggg	aga	576	
Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	Gly	Arg		
			180				185						190				
atg	tta	atc	cat	att	ggc	aga	acg	aaa	acg	ctg	gtt	agc	acc	gca	ggg	624	
Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	Ala	Gly		
		195				200						205					
gta	gag	aag	gca	ctt	agc	ctg	ggg	gta	act	aaa	ctg	gtc	gag	cga	tgg	672	
Val	Glu	Lys	Ala	Leu	Ser	Leu	Gly	Val	Thr	Lys	Leu	Val	Glu	Arg	Trp		
	210					215					220						
att	tcc	gtc	tct	ggg	gta	gct	gat	gat	cgc	aat	aac	tac	ctg	ttt	tgc	720	
Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	Phe	Cys		
	225				230					235					240		
cgg	gtc	aga	aaa	aat	ggg	gtt	gcc	gcg	cca	tct	gcc	acc	agc	cag	cta	768	
Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	Gln	Leu		
			245						250					255			
tca	act	cgc	gcc	ctg	gaa	ggg	att	ttt	gaa	gca	act	cat	cga	ttg	att	816	
Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	Leu	Ile		
			260					265					270				
tac	ggc	gct	aag	gat	gac	tct	ggg	cag	aga	tac	ctg	gcc	tgg	tct	gga	864	
Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	Ser	Gly		
	275					280						285					
cac	agt	gcc	cgt	gtc	gga	gcc	gcg	cga	gat	atg	gcc	cgc	gct	gga	gtt	912	
His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	Gly	Val		
	290				295						300						
tca	ata	ccg	gag	atc	atg	caa	gct	ggg	ggc	tgg	acc	aat	gta	aat	att	960	
Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	Asn	Ile		
	305				310					315					320		
gtc	atg	aac	tat	atc	cgt	aac	ctg	gat	agt	gaa	aca	ggg	gca	atg	gtg	1008	
Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	Met	Val		
			325						330					335			
cgc	ctg	ctg	gaa	gat	ggc	gat	tag									1032	
Arg	Leu	Leu	Glu	Asp	Gly	Asp	*										
			340														

<210> 59
 <211> 343
 <212> PRT

<213> Escherichia coli

<400> 59

```

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1      5      10      15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
20     25     30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Ser Val
35     40     45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
50     55     60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Tyr Leu Gln Ala
65     70     75     80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
85     90     95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
100    105    110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
115    120    125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
130    135    140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
145    150    155    160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
165    170    175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
180    185    190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
195    200    205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
210    215    220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
225    230    235    240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
245    250    255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
260    265    270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
275    280    285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
290    295    300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
305    310    315    320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
325    330    335
Arg Leu Leu Glu Asp Gly Asp
340

```

<210> 60

<211> 1272

<212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)...(1272)

<223> nucleotide sequence encoding Flip recombinase

<400> 60

```

atg cca caa ttt ggt ata tta tgt aaa aca cca cct aag gtg ctt gtt
Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1      5      10      15

cgt cag ttt gtg gaa agg ttt gaa aga cct tca ggt gag aaa ata gca
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
20     25     30

```

48

96

tta tgt gct gct gaa cta acc tat tta tgt tgg atg att aca cat aac Leu Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn 35 40 45	144
gga aca gca atc aag aga gcc aca ttc atg agc tat aat act atc ata Gly Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile 50 55 60	192
agc aat tcg ctg agt ttc gat att gtc aat aaa tca ctc cag ttt aaa Ser Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys 65 70 75 80	240
tac aag acg caa aaa gca aca att ctg gaa gcc tca tta aag aaa ttg Tyr Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu 85 90 95	288
att cct gct tgg gaa ttt aca att att cct tac tat gga caa aaa cat Ile Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His 100 105 110	336
caa tct gat atc act gat att gta agt agt ttg caa tta cag ttc gaa Gln Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu 115 120 125	384
tca tcg gaa gaa gca gat aag gga aat agc cac agt aaa aaa atg ctt Ser Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu 130 135 140	432
aaa gca ctt cta agt gag ggt gaa agc atc tgg gag atc act gag aaa Lys Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys 145 150 155 160	480
ata cta aat tcg ttt gag tat act tcg aga ttt aca aaa aca aaa act Ile Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr 165 170 175	528
tta tac caa ttc ctc ttc cta gct act ttc atc aat tgt gga aga ttc Leu Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe 180 185 190	576
agc gat att aag aac gtt gat ccg aaa tca ttt aaa tta gtc caa aat Ser Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn 195 200 205	624
aag tat ctg gga gta ata atc cag tgt tta gtg aca gag aca aag aca Lys Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Thr Glu Thr Lys Thr 210 215 220	672
agc gtt agt agg cac ata tac ttc ttt agc gca agg ggt agg atc gat Ser Val Ser Arg His Ile Tyr Phe Phe Ser Ala Arg Gly Arg Ile Asp 225 230 235 240	720
cca ctt gta tat ttg gat gaa ttt ttg agg aat tct gaa cca gtc cta Pro Leu Val Tyr Leu Asp Glu Phe Leu Arg Asn Ser Glu Pro Val Leu 245 250 255	768
aaa cga gta aat agg acc ggc aat tct tca agc aat aaa cag gaa tac Lys Arg Val Asn Arg Thr Gly Asn Ser Ser Asn Lys Gln Glu Tyr 260 265 270	816
caa tta tta aaa gat aac tta gtc aga tcg tac aat aaa gct ttg aag Gln Leu Leu Lys Asp Asn Leu Val Arg Ser Tyr Asn Lys Ala Leu Lys 275 280 285	864
aaa aat gcg cct tat tca atc ttt gct ata aaa aat ggc cca aaa tct Lys Asn Ala Pro Tyr Ser Ile Phe Ala Ile Lys Asn Gly Pro Lys Ser 290 295 300	912

cac att gga aga cat ttg atg acc tca ttt ctt tca atg aag ggc cta	960
His Ile Gly Arg His Leu Met Thr Ser Phe Leu Ser Met Lys Gly Leu	
305 310 315 320	
acg gag ttg act aat gtt gtg gga aat tgg agc gat aag cgt gct tct	1008
Thr Glu Leu Thr Asn Val Val Gly Asn Trp Ser Asp Lys Arg Ala Ser	
325 330 335	
gcc gtg gcc agg aca acg tat act cat cag ata aca gca ata cct gat	1056
Ala Val Ala Arg Thr Thr Tyr Thr His Gln Ile Thr Ala Ile Pro Asp	
340 345 350	
cac tac ttc gca cta gtt tct cgg tac tat gca tat gat cca ata tca	1104
His Tyr Phe Ala Leu Val Ser Arg Tyr Tyr Ala Tyr Asp Pro Ile Ser	
355 360 365	
aag gaa atg ata gca ttg aag gat gag act aat cca att gag gag tgg	1152
Lys Glu Met Ile Ala Leu Lys Asp Glu Thr Asn Pro Ile Glu Glu Trp	
370 375 380	
cag cat ata gaa cag cta aag ggt agt gct gaa gga agc ata cga tac	1200
Gln His Ile Glu Gln Leu Lys Gly Ser Ala Glu Gly Ser Ile Arg Tyr	
385 390 395 400	
ccc gca tgg aat ggg ata ata tca cag gag gta cta gac tac ctt tca	1248
Pro Ala Trp Asn Gly Ile Ile Ser Gln Glu Val Leu Asp Tyr Leu Ser	
405 410 415	
tcc tac ata aat aga cgc ata taa	1272
Ser Tyr Ile Asn Arg Arg Ile *	
420	

<210> 61
 <211> 422
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 61
 Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val Arg
 1 5 10 15
 Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala Leu
 20 25 30
 Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn Gly
 35 40 45
 Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile Ser
 50 55 60
 Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys Tyr
 65 70 75 80
 Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu Ile
 85 90 95
 Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His Gln
 100 105 110
 Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu Ser
 115 120 125
 Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu Lys
 130 135 140
 Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys Ile
 145 150 155 160
 Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr Leu
 165 170 175
 Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe Ser
 180 185 190
 Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn Lys
 195 200 205
 Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr Ser

210	215	220
Val Ser Arg His Ile Tyr	Phe Phe Ser Ala Arg	Gly Arg Ile Asp Pro
225	230	235
Leu Val Tyr Leu Asp Glu	Phe Leu Arg Asn Ser	Glu Pro Val Leu Lys
245	250	255
Arg Val Asn Arg Thr Gly	Asn Ser Ser Ser Asn	Lys Gln Glu Tyr Gln
260	265	270
Leu Leu Lys Asp Asn Leu	Val Arg Ser Tyr Asn	Lys Ala Leu Lys Lys
275	280	285
Asn Ala Pro Tyr Ser Ile	Phe Ala Ile Lys Asn	Gly Pro Lys Ser His
290	295	300
Ile Gly Arg His Leu Met	Thr Ser Phe Leu Ser	Met Lys Gly Leu Thr
305	310	315
Glu Leu Thr Asn Val Val	Gly Asn Trp Ser Asp	Lys Arg Ala Ser Ala
325	330	335
Val Ala Arg Thr Tyr Thr	His Gln Ile Thr Ala	Ile Pro Asp His
340	345	350
Tyr Phe Ala Leu Val Ser	Arg Tyr Tyr Ala Tyr	Asp Pro Ile Ser Lys
355	360	365
Glu Met Ile Ala Leu Lys	Asp Glu Thr Asn Pro	Ile Glu Glu Trp Gln
370	375	380
His Ile Glu Gln Leu Lys	Gly Ser Ala Glu Gly	Ser Ile Arg Tyr Pro
385	390	395
Ala Trp Asn Gly Ile Ile	Ser Gln Glu Val Leu	Asp Tyr Leu Ser Ser
405	410	415
Tyr Ile Asn Arg Arg Ile		
420		

<210> 62
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR2

<400> 62
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttc 48

<210> 63
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> IR1

<400> 63
 gaagttccta tactttctag agaataggaa ctccggaata ggaacttc 48

<210> 64
 <211> 66
 <212> DNA
 <213> Bacteriophage mu

<220>
 <221> CDS
 <222> (1)...(66)
 <223> nucleotide sequence encoding GIN recombinase

<400> 64
 tca act ctg tat aaa aaa cac ccc gcg aaa cga gcg cat ata gaa aac 48
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
 1 5 10 15

gac gat cga atc aat taa 66
 Asp Asp Arg Ile Asn *

20

<210> 65
<211> 21
<212> PRT
<213> bacteriophage mu

<400> 65
Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn
1 5 10 15
Asp Asp Arg Ile Asn
20

<210> 66
<211> 69
<212> DNA
<213> Bacteriophage mu

<220>
<221> CDS
<222> (1)...(69)
<223> nucleotide sequence encoding Gin recombinase

<400> 66
tat aaa aaa cat ccc gcg aaa cga acg cat ata gaa aac gac gat cga 48
Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
1 5 10 15
atc aat caa atc gat cgg taa 69
Ile Asn Gln Ile Asp Arg *
20

<210> 67
<211> 22
<212> PRT
<213> bacteriophage mu

<220>
<223> Gin recombinase of bacteriophage mu

<400> 67
Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg
1 5 10 15
Ile Asn Gln Ile Asp Arg
20

<210> 68
<211> 555
<212> DNA
<213> Escherichia coli

<220>
<221> CDS
<222> (1)...(555)
<223> nucleotide sequence encoding PIN recombinase

<400> 68
atg ctt att ggc tat gta cgc gta tca aca aat gac cag aac aca gat 48
Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15
cta caa cgt aat gcg ctg aac tgt gca gga tgc gag ctg att ttt gaa 96
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu
20 25 30

gac aag ata agc ggc aca aag tcc gaa agg ccg gga ctg aaa aaa ctg	144
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
ctc agg aca tta tcg gca ggt gac act ctg gtt gtc tgg aag ctg gat	192
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
cgg ctg ggg cgt agt atg cgg cat ctt gtc gtg ctg gtg gag gag ttg	240
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
cgc gaa cga ggc atc aac ttt cgt agt ctg acg gat tca att gat acc	288
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
agc aca cca atg gga cgc ttt ttc ttt cat gtg atg ggt gcc ctg gct	336
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
gaa atg gag cgt gaa ctg att gtt gaa cga aca aaa gct gga ctg gaa	384
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
act gct cgt gca cag gga cga att ggt gga cgt cgt ccc aaa ctt aca	432
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Pro Lys Leu Thr	
130 135 140	
cca gaa caa tgg gca caa gct gga cga tta att gca gca gga act cct	480
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	
cgc cag aag gtg gcg att atc tat gat gtt ggt gtg tca act ttg tat	528
Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr	
165 170 175	
aag agg ttt cct gca ggg gat aaa taa	555
Lys Arg Phe Pro Ala Gly Asp Lys *	
180	

<210> 69

<211> 184

<212> PRT

<213> Escherichia coli

<400> 69

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp	
1 5 10 15	
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu	
20 25 30	
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	

Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr
 165 170 175
 Lys Arg Phe Pro Ala Gly Asp Lys
 180

<210> 70
 <211> 4778
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pcx plasmid

<400> 70
 gtcgacattg attattgact agttattaat agtaaatcaat tacgggggtca ttagttcata 60
 gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc 120
 ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
 ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtac 240
 atcaagtgtat tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
 cctgggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360
 tattagtcat cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
 atctccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
 gcgatggggg cggggggggg gggggcgccg gccaggcggg gcggggcggg gcgagggggc 540
 gggcgggggc aggcggagag gtgcccgggc agccaatcag agcgggcgcc tccgaaagtt 600
 tccttttatg gcgagggcgc ggcggcgccg gccctataaa aagcgaagcg cgcggcgggc 660
 gggagtcgct gcgttgccct cgcgccgtgc cccgctccgc gccgcctcgc gccgcccgc 720
 ccgctctga ctgaccgct tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780
 gggctgtaat tagcgcttg tttaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840
 ccttaaaggc ctccgggagg gccctttgtg cgggggggag cggctcgggg ggtgctgctg 900
 tgtgtgtgtg cgtggggagc gcccgctgct gcccgcgctg cccggcgctg gtgagcgctg 960
 cggcgcgccg cgggggctt gtgctgctgc aggggagcgc gggcgggggc 1020
 ggtgccccgc ggtgccccgc ggtgccccgc ggaacaaagg ctgctgctgc ggtgtgtgctg 1080
 tggggggggg agcagggggg gtggggcgcc cggctcgggc gtaaccccc cctgcacccc 1140
 cctccccgag ttgctgagca cggcccggtc tcgggtgcgg ggctccgtgc ggggctgtgg 1200
 gcggggctcg ccgtgcccgc cgggggggtg cggcaggtgg ggggtgcccgc cggggcgggg 1260
 ccgctccggc ccggggaggc ctccggggag gggcgcgccg gccccggagc gccggcggtc 1320
 gtcgagggcg ggcgagccgc agccattgcc tggcggagcc gaaatctggg aggcgcgcgc gcacccccctc 1440
 gacttccctt gtcccaaatc cggcgccgct cggggcgagg ggcaggaagg aaatggcgcg ggagggcctt 1500
 tagcggcgcc gggcggaagc gtgcccgcgc tccccttctc catctccagc ctccgggctg ccgagggggg 1560
 cgtgctgcgc cgcgcgcgc cggggcgagg cgggggtcgg cttctggcgt gtgacggcg 1620
 acgggtgcct tcggggggga catgttcatg cttcttctt tttctacag ctctgggca 1680
 gctctagagc ctctgctaac tctcatcatt ttggcaaga attcactcct cagggtgcagg 1740
 acgtgctggt tggtgtgctg gctggtgtgg gctggtgctt ccaatgcact taccactgag 1800
 ctgcctatca gaaggtggtg aattttggg atattgccc attttcattt agggcacaatc 1860
 atctttttcc ctctgcaaaa attttcattt attttgttt agggcacaatc 1920
 ctggctaata aaggaaattt agggcacaatc attttgttt agggcacaatc 1980
 cactcggaag gacatatggg agggcacaatc attttgttt agggcacaatc 2040
 agagtgtggc aacatatgoc atattgctgg atattgccc attttgttt 2100
 catcagtata tgaacagacc cctgctgtgc attttgttt agggcacaatc 2160
 tgaggttaga ttttttttat attttgttt agggcacaatc attttgttt 2220
 ttttccttac atgttttact agccagattt aagatccctc gaaattgtta 2280
 gctgtccctc tttctctatg aagatccctc gaaattgtta tccgctcaca attccacaca 2340
 gtcataagctg tttctctgtg aagatccctc gaaattgtta tccgctcaca attccacaca 2400
 cggaagcata aagtgtaaag cctgggggtg ctaatgagtg agctaactca cattaattgc 2460
 gttgcgctca ctgcccgtt tocagtcgg ctaatgagtg agctaactca cattaattgc 2520
 aattagtcag caaccatagt cccgccccta ctaattttt ttatttatgc agaggccgag 2580
 agttccgccc attctccgct ccatggctga tagtgaggag gcttttttgg aggcctaggc 2640
 gccgcctcgg cctctgagct attccagaag cttataatgg ttacaaataa agcaatagca 2700
 ttttgcaaaa agctaacttg tttattgcag cactgcattc tagttgtggt ttgtccaaac 2760
 tcacaaattt cacaaataaa gcattttttt gctgcatcaa tgaatcgccc aacgcgcggg 2820
 tcatcaatgt atcttatcat gtctggatcc gctgcatcaa tgaatcgccc aacgcgcggg 2880
 gagagcggt ttgctgattg ggcgctcttc cgttctctcg ctactgact cgtctgcttc 2940
 ggtcgttcgg ctgcccgcag cggatcagc gaaagaacat gtagcagaaa agggcaggaa 3000
 agaatcaggg gataacgcag tggcggtttt aaacccgaca ggactataaa gataccaggc 3060
 ccgtaaaaag gcgcgattgc agaggtggcg aaacccgaca ggactataaa gataccaggc 3120
 caaaaatcga cgctcaagtc tgcgtgcgtc tcctgttccg accctgccc 3180
 gtttccccct ggaagctccc

cctgtccgcc	tttctccctt	cggaagcgt	ggcgctttct	caatgctcac	gctgtaggta	3240
tctcagttcg	gtgtaggtcg	tctgtccaa	gctgggctgt	gtgcacgaac	ccccggttca	3300
gcccgaccgc	tgccgcttat	ccggtaacta	tcgtcttgag	tccaacccgg	taagacacga	3360
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	3420
tgctacagag	ttcttgaagt	ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	3480
tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	3540
caaacaaacc	accgctggta	gcggtggttt	ttttgtttgc	aagcagcaga	ttacgcgcag	3600
aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtggaa	3660
cgaaaactca	cgtaaaggga	ttttggatcat	gagattatca	aaaaggatct	tcacctagat	3720
ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	3780
tgacagttac	caatgcttaa	tcagtggagg	acctatctca	gcgatctgtc	tatttcgttc	3840
atccatagtt	gcctgactcc	cgtgctgta	gataactacg	atcgggagg	gcttaccatc	3900
tgcccccagt	gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atttatcagc	3960
aataaaccag	ccagccggaa	gggcccagcg	cagaagtggg	cctgcaactt	tatccgcctc	4020
catccagtc	attaatgtt	gcccgggaagc	tagagtaagt	agttccgccg	ttaatagttt	4080
gcgcaacggt	tgtgccattg	ctacaggcat	cgtgggttca	cgctcgtcgt	ttggtagtgc	4140
ttcattcagc	tcgggttccc	aacgatcaag	gagagttaca	tgatccccca	tggtgtgcaa	4200
aaaagcggtt	agctccttcg	gtcctccgat	cgttgctcaga	agtaagttgg	ccgcagtgtt	4260
atcactcatg	gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	4320
cttttctgtg	actgggtgag	actcaaccaa	gtcattctga	gaatagtgtg	tgccggcgacc	4380
gagttgctct	tgcccggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	4440
agtgtcatc	attggaaaaac	gttcttcggg	gcgaaaaactc	tcaaggatct	taccgctggt	4500
gagatccagt	tcgatgtaac	ccactcgtgc	accacaatga	tcttcagcat	cttttacttt	4560
caccagcgtt	tctgggtgag	caaaaaacagg	aaggcaaaat	gccgcaaaaa	agggaaataag	4620
ggcgacacgg	aaatgttgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	4680
tcagggttat	tgtctcatga	gcggatacat	atttgaatgt	atttagaaaa	ataaacaaat	4740
aggggttccg	cgcacatttc	ccgaaaaagt	gccacctg			4778

<210> 71

<211> 5510

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFP plasmid

<400> 71

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataatat	ggagttccgc	gttacataac	ttacggtaaa	tgcccccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgcacaatag	180
ggacttttcca	ttgacgtcaa	tggtgggact	atttaccggt	aactgcccac	ttggcagtag	240
atcaagtgtg	tcataatgcca	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccc	300
cctggcattg	tgcccagtag	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcac	ctctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tattttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccaggcggg	gcggggcggg	gcgagggggc	540
ggggcgggcg	aggcgagag	gtgcggcgcg	agccaatcag	agcgggcgcg	tccgaaagt	600
tccttttatg	gcgaggcgcg	ggcgggcgcg	gccctataaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgctccgc	gccgcctcgc	gccgccccgc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgctgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggtcgggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagg	gccgcgtgcg	gcccgcgctg	ccggcgggct	gtgagcgctg	960
cggggcgggc	gcggggcctt	gtgcgtcccg	cgtgtgcgcg	aggggagcgc	ggccgggggc	1020
ggtgccccgc	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	ggtgtgtgcg	1080
tggggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaacccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggct	tcgggtgcgg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	ccgtgccggg	cggggggtgg	cggcaggtgg	gggtgcgggg	cgggggcggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gcccgggagc	gccggcgggc	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgcgag	1380
gaacttcttt	gtcccaaata	tggcggagcc	gaaatctggg	aggcgccgcc	gcacccccct	1440
tagcgccgcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatggggcg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cgggggttcgg	cttctggcgt	gtgacggcg	1620
gctctagagc	ctctgctaac	catgttcatt	ccttctcttt	tttctacag	ctcctgggca	1680
acgtgctggg	tgttgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atggtaggca	1740
agggcgaggga	gctgttcacc	gggggtggtg	ccatcctggt	cgaactggac	ggcgacgtaa	1800

acggccacaa	gttcagcgtg	tccggcgagg	gcgaggggga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgaccta	cgcggtgcag	tgcttcagcc	gctaccccga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gcccagggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tgcccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccc	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcgac	cctgagcaaa	gaccccaacg	agaagcgcgga	tcacatgggtc	ctgctggagt	2400
tcgtgaccgc	cgccgggatc	actctcggca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtgg	tggtcggtgt	ggccaatgcc	ctggctcaca	2520
aataaccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccccttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tggttgaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcatttataa	catcagaatg	2700
agtatttggt	ttagagtgtg	gcaacatatg	ccatatgctg	gctgccatga	acaaagggtg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattccctt	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atattttgtt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttcctt	acatgtttta	ctagccagat	ttttcctcct	ctcctgacta	2940
ctcccagtca	tagctgtccc	tcctctctta	tgaagatccc	tcgacctgca	gcccgaagctt	3000
ggcgtaaatca	tggtcatagc	tgtttcctgt	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gcccgaagca	taaagtgtaa	agcctggggg	gcctaagtga	tgagctaact	3120
cacattaatt	gcgttgcgct	cactgccgcg	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcctc	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgccc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatgggt	gactaatttt	ttttatttat	3300
gcagagggcg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggccctag	gcttttgcaa	aaagctaaat	tgtttatggc	agcttataat	ggttacaaat	3420
aaagcaatag	catcacaaat	ttcacaaata	aagcattttt	ttcactgcat	tctagttgtg	3480
gtttgtccaa	actcatcaat	gtatcttatc	atgtctggat	ccgctgcatt	aatgaatcgg	3540
ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctct	tccgcttcct	cgctcactga	3600
ctcgctgcgc	tcggtcgttc	ggctgcccgc	agcgggtatc	gctcactcaa	aggcggtaat	3660
acgggttatcc	acagaatcag	gggataaacgc	aggaaagaac	atgtgagcaa	aaggccagca	3720
aaaggccagg	aaccgtaaaa	aggccgcggt	gctggcggtt	ttccataggc	tccgcccccc	3780
tgaccgagcat	cacaaaaatc	gacgctcaag	tcagagggtg	cgaaaccoga	caggactata	3840
aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	tctcctgttc	cgaccctgcc	3900
gcttacccga	tacctgtccg	cctttctccc	ttcgggaagc	gtggcgcttt	ctcaatgctc	3960
acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	aagctggggt	gtgtgcacga	4020
accccccggt	cagccccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaaccc	4080
ggtaagacac	gacttatcgc	cactggcagc	agccactggg	aacaggatta	gcagagcgag	4140
gtatgtaggc	gggtgctacag	agttcttgaa	gtggtggcct	aactacggct	acactagaag	4200
gacagtattt	ctctgctgaa	ccactgctgg	gccagttaac	ttcggaaaaa	gagttggtag	4260
ctcttgatcc	ggcaaacaaa	ccacgcctgg	tagcggtggt	ttttttgttt	gcaagcagca	4320
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atcttttcta	cgggggtctga	4380
cgctcagtg	aacgaaaaact	cacgttaagg	gattttgggt	atgagattat	caaaaaggat	4440
cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	tcaatctaaa	gtatatatga	4500
gtaaacttgg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4560
tctatttcgt	tcattccatag	ttgcctgact	ccccgtcgtg	tagataacta	cgatacggga	4620
gggcttaacca	tctggcccca	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4680
agattttatca	gcaataaacc	agccagcccg	aagggccgag	cgcagaagtg	gtcctgcaac	4740
tttatccgcc	tccatccagt	ctattaattg	ttgcggggaa	gctagagtaa	gtagttcgcc	4800
agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	atcgtgggtg	cacgctcgtc	4860
gtttggtagt	gcttcattca	gctccgggtc	ccaacgatca	aggcgagtta	catgatcccc	4920
catggttggtc	aaaaaagcgg	ttagctcctt	cggtcctccg	atcgttgctc	gaagtaagtt	4980
ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgcc	5040
atccgtaaga	tgcttttctg	tgactgggtg	gtactcaacc	aagtcattct	gagaatagtg	5100
tatggggcga	ccgagttgct	cttgcccggc	gtcaatacgg	gataataccg	cgccacatag	5160
cagaacttta	aaagtgtcca	tcattggaaa	acgttcttcg	gggcgaaaaa	tctcaaggat	5220
cttacccgctg	ttgagatcca	gttcgatgta	acccactcgt	gcacccaact	gatcttcagc	5280
atctttttact	ttcaccagcg	tttctgggtg	agcaaaaaa	ggaaggcaaa	atgccgcaaa	5340
aaagggaata	agggcgacac	ggaaatgttg	aatactcata	ctcttccttt	ttcaatatta	5400
ttgaagcatt	tatcagggtt	attgtctcat	gagcgggata	atatttgaat	gtatttagaa	5460
aaataaaca	ataggggttc	cgcgcacatt	tcccgcgaaa	gtgccacctg		5510

<210> 72
 <211> 282
 <212> DNA
 <213> Artificial Sequence

<220>
<223> attp

<400> 72
ccttgcgcta atgctctggt acaggctcact aataccatct aagtagttga ttcatagtga 60
ctgcataatgt tgtgtttttac agtattatgt agtctgtttt ttatgcaaaa tctaatttaa 120
tatattgata tttatatcat tttacgtttc tcgttcagct tttttatact aagttggcat 180
tataaaaaag cattgcttat caatttggtg caacgaacag gtcactatca gtcaaaataa 240
aatcattatt tgatttcaat tttgtcccac tccctgcctc tg 282

<210> 73
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 73
ggccccgtaa tgcagaagaa 20

<210> 74
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 74
ggtttaaagt gcgctcctcc aagaacgtca tc 32

<210> 75
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 75
agatctagag ccgccgctac aggaacaggt ggtggcggcc 40

<210> 76
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5PacSV40

<400> 76
ctgttaatta actgtggaat gtgtgtcagt taggggtg 37

<210> 77
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer Antisense Zeo

<400> 77
tgaacagggt cacgtcgtcc 20

<210> 78
<211> 24

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5' HETS

<400> 78
gggccgaaac gatctcaacc tatt 24

<210> 79
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3' HETS

<400> 79
cgcagcggcc ctcctactc 19

<210> 80
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 5BSD

<400> 80
accatgaaaa catttaacat ttctcaaca 29

<210> 81
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer SV40polyA

<400> 81
tttatttggtg aaatttggtga tgctattgc 29

<210> 82
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer 3BSP

<400> 82
ttaatttcgg gtatatattga gtgga 25

<210> 83
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer EPO5XBA

<400> 83
tatctagaat gggggtgcac gaatgtcctg cc 32

<210> 84
<211> 32
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer EPO3SBI

<400> 84

tacgtacgtc atctgtcccc tgtcctgcag gc

32

<210> 85

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEP03BSI

<400> 85

cgtacgtcat ctgtccccctg tctctgca

27

<210> 86

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEP05XBA

<400> 86

tctagaatgg ggggtgcacgg tgagtact

28

<210> 87

<211> 4862

<212> DNA

<213> Artificial Sequence

<220>

<223> pD2eGFP-1N plasmid from Clontech

<400> 87

tagttattaa	tagtaatcaa	ttacgggggtc	attagttcat	agcccatata	tgaggttccg	60
cgttacataa	cttacggtaa	atggcccgcgc	tggttgaccg	cccaacgacc	cccggccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacggt	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tgcctattac	360
catgggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	ctccaccoca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	540
acggtgggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacgggtacc	gcggggcccg	660
gatccaccgg	tcgccaccat	ggtgagcaag	ggcgaggagc	tgttcaccgg	gggtggtgcc	720
atcctggctcg	agctggacgg	cgacgtaaac	ggccacaagt	tcagcgtgtc	cggcgagggc	780
gagggcgatg	ccacctacgg	caagctgacc	ctgaagttca	tctgcaccac	cggcaagctg	840
cccggtgccct	ggcccaccct	cgtgaccacc	ctgacctacg	gcgtgcagtg	cttcagccgc	900
taccccgacc	acatgaagca	gcacgacttc	ttcaagtccg	ccatgcccg	aggctacgtc	960
caggagcgca	ccatcttctt	caaggacgac	ggcaactaca	agaccgcgc	cgaggtgaag	1020
ttcgaggggcg	acaccctggg	gaaccgcatc	gagctgaagg	gcatcgactt	caaggaggac	1080
ggcaacatcc	tgggggcaca	gctggagtac	aactacaaca	gccacaacgt	ctatatcatg	1140
gccgacaagc	agaagaacgg	catcaagggtg	aacttcaaga	tccgccacaa	catcgaggac	1200
ggcagcgtgc	agctcgccga	ccactaccag	cagaacaccc	ccatcgccga	cggccccgtg	1260
ctgctgcccg	acaaccacta	cctgagcacc	cagtcgcggc	tgagcaaaga	ccccaacgag	1320
aagcgcgatc	acatggtcct	gctggagtgc	gtgaccgccc	ccgggatcac	tctcggcatg	1380
gacgagctgt	acaagaagct	tagccatggc	ttcccgcggg	aggtggagga	gcaggatgat	1440
ggcacgctgc	ccatgtcttg	tgcccaggag	agcgggatgg	accgtcacc	tgcagcctgt	1500
gcttctgcta	ggatcaatgt	gtagatgcgc	ggccgcgact	ctagatcata	atcagccata	1560
ccacatttgt	agaggtttta	cttgctttta	aaaacctccc	acacctccc	ctgaacctga	1620
aacataaaat	gaatgcaatt	gttgttggtta	acttgtttat	tgcagcttat	aatgggttaca	1680

aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttcaactg	cattctagtt	1740
gtgggtttgtc	caaactcacc	aatgtatctt	aaggcgtaaa	ttgtaagcgt	taatatTTTTg	1800
ttaaaatttcg	cgttaaattt	ttgttaaatac	agctcatttt	ttaaccaata	ggccgaaatc	1860
ggcaaaatcc	cttataaaatc	aaaagaatag	accgagatag	ggttgagtgt	tgttccagtt	1920
tggaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaagggcg	aaaaaccgtc	1980
tatcagggcg	atggcccact	acgtgaacca	tcaccctaata	caagtttttt	ggggtcgagg	2040
tgccgtaaaag	cactaaatcg	gaaccctaaa	gggagccccc	gatttagagc	ttgacgggga	2100
aagccggcgga	acgtggcgag	aaaggaaggg	aagaaagcga	aaggagcggg	cgctagggcg	2160
ctggcaagtg	tagcgggtcac	gctgcgcgta	accaccacac	ccgccgcgct	taatgcgccg	2220
ctacagggcg	cgtcagggtg	cactttttcgg	ggaaatgtgc	gcggaacccc	tatttgttta	2280
tttttctaaa	tacattcaaaa	tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	2340
caataatatt	gaaaaaggaa	gagtcctgag	gcggaaagaa	ccagctgtgg	aatgtgtgtc	2400
agttagggtg	tggaaggtcc	ccaggctccc	cagcaggcag	aagtatgcaa	agcatgcac	2460
tcaattagtc	agcaaccagg	tgtggaaagt	ccccaggctc	cccagcaggc	agaagtatgc	2520
aaagcatgca	tctcaattag	tcagcaacca	tagtcccgcg	cctaactccg	cccatcccg	2580
ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	ctgactaatt	ttttttattt	2640
atgcagaggc	cgaggccgccc	tcggcctctg	agctattcca	gaagtagtga	ggaggctttt	2700
ttggaggcct	aggctttttgc	aaagatcgat	caagagacag	gatgaggatc	gtttcgcatg	2760
attgaacaag	atggattgca	cgaggttct	ccggccgctt	gggtggagag	gctattcggc	2820
tatgactggg	cacaacagac	aatcggtcgc	tctgatgccg	ccgtgttccg	gctgtcagcg	2880
caggggcccgc	cggttctttt	tgtcaagacc	gacctgtccg	gtgccctgaa	tgaactgcaa	2940
gacgaggcag	cgcggtatc	gtggctggcc	acgacgggcg	ttccttgccg	agctgtgctc	3000
gacgttgtca	ctcaagcggg	aagggaactg	ctgctattgg	gcgaagtgcc	ggggcaggat	3060
ctcctgtcat	ctcagccttg	tcctgcccag	aaagtatcca	tcattggctga	tgcaatgcgg	3120
cggtctgcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	acatcgcatc	3180
gagcgagcac	gtactcggat	ggaagccggg	cttgtcgatc	aggatgatct	ggacgaagag	3240
catcagggcg	tcgcgccagc	cgaactgttc	gccaggctca	aggcgagcat	gcccagcgcc	3300
gaggatctcg	tcgtgaccca	tgggcatgcc	tgcttgccga	atatcatggt	ggaaaatggc	3360
cgcttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaaccgcta	tcaggacata	3420
gcgttggtca	cccgatgat	tgctgaagag	cttgccggcg	aatgggctga	ccgcttcctc	3480
gtgctttacg	gtatcgccgc	tcctgatctg	cagcgcatcg	ccttctatcg	ccttcttgac	3540
gagttcttct	gagcgggact	ctgggttctg	aaatgaccga	ccaagcgacg	cccaaccctg	3600
catcacgaga	tttcgattcc	accgcccgtc	tctatgaaag	gttgggcttc	ggaatcgttt	3660
tccgggacgc	cggtctggatg	atcctccagc	gcggggatct	catgctggag	ttcttcgccc	3720
accctagggg	gaggctaact	gaaacacgga	aggagacaat	accggaagga	acccgcgcta	3780
tgacggcaat	aaaaagacag	aataaaaacg	acggtgttgg	gtcgtttgtt	cataaacgcg	3840
gggttcggtc	ccagggtcgg	cactctgtcg	ataccccacc	gagaccccat	tggggccaat	3900
acgcccgcgt	ttcttccctt	tcctccaccc	accccccaag	ttcgggtgaa	ggcccagggc	3960
tcgcagccaa	cgtcggggcg	gcaggccctg	ccatagccctc	aggttactca	tatatacttt	4020
agattgattt	aaaacttcatt	ttttaattta	aaaggatcta	ggtgaagatc	ctttttgata	4080
atctcatgac	caaaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	gaccccgtag	4140
aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	cgtaatctgc	tgcttgcaaa	4200
caaaaaaac	accgctacca	gcggtgggtt	gtttgcccga	tcaagagcta	ccaactcttt	4260
ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	4320
cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	gctctgctaa	4380
tcctgttacc	agtggctgct	gccagtgccg	ataagtcgtg	tccttaccggg	ttggactcaa	4440
gacgatagtt	accggataag	gcgcagcggg	cgggctgaac	gggggggttcg	tgcacacagc	4500
ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	ctatgagaaa	4560
gcgccacgct	tcccgaaggg	agaaaggcgg	acaggtatcc	ggtaagcggc	agggtcggaa	4620
caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	agtctgtcgc	4680
ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	ggcgaggacc	4740
tatggaaaaa	cgccagcaac	gcggcctttt	tacggttctc	ggccttttgc	tggccttttg	4800
ctcacatggt	ctttcctgcg	ttatccctcg	attctgtgga	taaccgtatt	accgccatgc	4860
at						4862

<210> 88

<211> 5192

<212> DNA

<213> Artificial Sequence

<220>

<223> pIRESpuro2 plasmid from Clontech

<400> 88

gacggatccg	gagatctccc	gatccctat	ggctgactct	cagtacaatc	tgctctgatg	60
ccgcatagtt	aagccagtat	ctgctccctg	cttgtgtgtt	ggaggctcgt	gagtagtgcg	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgaccga	caattgcatg	aagaatctgc	180

ttaggggttag	gcggttttgcg	ctgcttcgcg	atgtacgggc	cagatatagc	cggttgacatt	240
gattattgac	tagttatttaa	tagtaatacaa	ttacgggggc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	360
cccgccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tattttacggg	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaattggccc	gcctggcatt	540
atgccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catgggtgatg	cggttttggc	agtacatcaa	tgggctgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaattggcg	780
gtaggcggtg	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttgggtacc	900
gagctcggat	cgatatctgc	ggcctagcta	gcgcttaagg	cctgttaacc	ggctcgtagct	960
ctccggattc	gaattcggat	ccgcgccgcg	atagataact	gatccagtgt	gctggaatta	1020
attcgctgtc	tgcgagggcc	agctgttggg	gtgagtactc	cctctcaaaa	gcgggcatga	1080
cttctgcgct	aagattgtca	gtttccaaaa	acgaggagga	tttgatattc	acctggcccg	1140
cggtgatgcc	tttgaggggtg	gccgcgtcca	tctggtcaga	aaagacaatc	ttttgttgt	1200
caagcttgag	gtgtggcagg	cttgagatct	ggccatacac	ttgagtga	atgacatcca	1260
ctttgccttt	ctctccacag	gtgtccactc	ccagggtcaa	ctgcaggctc	agcatgcac	1320
tagggcgccc	ctctccgccc	ctctccctcc	cccccccta	acgttactgg	ccgaagccgc	1380
ttggaataag	gccggtgtgc	gtttgtctat	atgtgatttt	ccaccatatt	gccgtctttt	1440
ggcaatgtga	ggggccggaa	acctggccct	gtcttcttga	cgagcattcc	taggggtctt	1500
tccctctctc	cccaagggaat	gcaaggctctg	ttgaatgtcg	tgaagggaagc	agttcctctg	1560
gaagcttctt	gaagacaaac	aacgtctgtg	gcgacccttt	gcaggcagcg	gaacccccc	1620
cctggcgaca	ggtgcctctg	cgggccaaaag	ccagctgtat	aagatacacc	tgcaaaaggcg	1680
gcacaacccc	agtgccacgt	tgtgagttgg	atagttgtgg	aaagagtcaa	atggctctcc	1740
tcaagcgatat	tcaacaaggg	gctgaaggat	gcccagaagg	tacccattg	tatgggatct	1800
gatctggggc	ctcggtgcac	atgctttaca	tgtgtttagt	cgagggttaaa	aaaacgtcta	1860
ggccccccga	accacgggga	cgtggttttc	ctttgaaaaa	cacgatgata	agcttgccac	1920
aacccacaag	gagacgacct	tccatgaccg	agtacaagcc	cacgggtgcg	ctcgccaccc	1980
gcgacgacgt	cccccgggcc	gtacgcaccc	tcgcccgcgc	gttcgcccac	taccccgcca	2040
cgcgccacac	cgctcgacccg	gaccgccaca	tcgagcgggt	caccgagctg	caagaactct	2100
tcctcacgcg	cgctcgggctc	gacatcgcca	aggtgtgggt	cgcggacgac	ggcgcccgcg	2160
tggcggtctg	gaccacgccc	gagagcgctg	aagcgggggc	ggtgttcgcc	gagatcggcc	2220
cgcgcatggc	cgagttgagc	ggttcccggc	tggccgcgca	gcaacagatg	gaaggcctcc	2280
tggcgccgca	ccggcccaag	gagccccgct	ggttctctgg	caccgtcggc	gtctcgcccg	2340
accaccaggg	cgagggtctg	ggcagcgccg	tcgtgtctcc	cggagtggag	gcggccgagc	2400
gcgcccgggt	gcccgccttc	ctggagacct	ccgcgccccg	caacctcccc	ttctacgagc	2460
ggctcggctt	caccgtcacc	gccgacgtcg	agtgcccga	ggaccgcgcg	acctgggtga	2520
tgacccgcaa	gcccgggtgc	tgacgcccgc	cccacgaccc	gcagcgcccc	accgaaaagg	2580
gcgcacgacc	ccatggctcc	gaccgaagcc	gaccggggcg	gccccgcgca	ccccgcaccc	2640
gcccccgagg	cccaccgact	ctagagctcg	ctgatcagcc	tcgactgtgc	cttctagtgt	2700
ccagccatct	gttgtttgccc	cctcccccg	gccttctctg	accttgaag	gtgccaactc	2760
cactgtcctt	tcctaataaa	atgaggaaat	tgctatcgat	tgtctgagta	ggtgtcattc	2820
tattctgggg	ggtgggggtg	ggcaggacag	caagggggag	gatttgggaag	acaatagcag	2880
gcatgctggg	gatgcgggtg	gctctatggc	ttctgaggcg	gaaagaacca	gctggggctc	2940
gagtgacctc	tagttgtggg	ttgtccaaac	tcactcaatgt	atcttatcat	gtctgtatac	3000
cgtcgacctc	tagctagagc	ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	3060
gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	aaagcctggg	3120
gtgcctaatt	agttagctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	gctttccagt	3180
cgggaaacct	gtcgtgccag	ctgcattaat	gaatcggcc	acgcgcgggg	agaggcggtt	3240
tgcgtattgg	gcgctcttcc	gcttctctgc	tcactgactc	gctgcgctcg	gtcgttcggc	3300
tgcggcgagc	ggtatcagct	cactcaaagg	cggtaatcac	ggtatccaca	gaatcagggg	3360
ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	ggccagggaac	cgtaaaaagg	3420
ccgcttgct	ggcggttttc	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	3480
gctcaagtca	gaggtggcga	aacccgacag	gactataaag	ataccaggcg	tttccccctg	3540
gaagctccct	cgtgcgctct	cctgttccga	ccctgcccgt	taccggatac	ctgtccgct	3600
ttctcccttc	gggaagcggt	gcgctttctc	aatgctcacg	ctgtagggtat	ctcagttcgg	3660
tgtaggtcgt	tcgctccaag	ctgggctgtg	tgacgaacc	ccccgttcag	cccgaaccgt	3720
gcgccttatc	cggtaactat	cgtcttgagt	ccaacccggg	aagacacgac	ttatcgccac	3780
tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	3840
tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggg	atctgcgctc	3900
tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	3960
ccgctggtag	cgggtggtttt	ttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	4020
ctcaagaaga	tcctttgatc	tttctacgg	ggcttgacgc	tcagtggaa	gaaaactcac	4080
gttaagggat	tttgggtcatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaatt	4140
aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttggctc	gacagttacc	4200

aatgcttaat	cagtggaggca	cctatctcag	cgatctgtct	atttcggttca	tccatagttg	4260
cctgactccc	cgctcgtgtag	ataactacga	tacgggaggg	cttaccatct	ggccccagtg	4320
ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	ataaaccagc	4380
cagccggaag	ggccgagcgc	agaagtgggtc	ctgcaacttt	atccgcctcc	atccagttcta	4440
tttaattgttg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	cgcaacgttg	4500
ttgccattgc	tacaggcatc	gtgggtgtcac	gctcgtcgtt	tggtatggct	tcattcagct	4560
ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	aaagcgggta	4620
gctccttcgg	tcctccgcatc	gttgtcagaa	gtaagttggc	cgcagtggtta	tcactcatgg	4680
ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	ttttctgtga	4740
ctgggtgagta	ctcaaccaag	tcattctgag	aatagtgtat	ggggcgaccg	agttgctctt	4800
gcccggcgctc	aatacgggat	aataccgcgc	cacatagcag	aactttaaaa	gtgctcatca	4860
ttggaaaacg	ttcttcgggg	cgaaaactct	caaggatctt	accgctgttg	agatccagtt	4920
cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	accagcggtt	4980
ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataaag	gcgacacgga	5040
aatgttgaat	actcatactc	ttcctttttt	aatattattg	aagcatttat	cagggttatt	5100
gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	ggggttcgcg	5160
gcacatttcc	ccgaaaagtg	ccacctgacg	tc			5192

<210> 89

<211> 11182

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg1 Plasmid

<400> 89

catgccaaac	acaggggttcc	cctcgggcatc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggcttctga	cggttcagtgc	agccgtcttc	tgaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgccgcgcc	tgcccttttc	ctggcggtttt	cttgtcgcgt	180
gttttagtgc	cataaagtag	aatacttgcg	actagaaccg	gagacattac	gccatgaaca	240
agagcgcgc	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggccgaactg	cacgcggcgc	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccgggactgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tgcccccag	caccgcgcac	ctactggaca	480
ttgccgagcg	cacccaggag	gcccggcgcg	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	gccggccggc	cgcatgggtg	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgcaaccgga	gcgggcgcga	ggccgcgaag	gcccgaggcg	660
tgaagtttgg	ccccgcct	accctcacc	cggcacagat	cgcgacgccc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaagagg	cggctgcact	gcttggcgtg	catcgctcga	780
ccctgtaccg	cgactttag	cgacgcgagg	aatgtacgcc	caccgaggcc	aggcggcgcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	ccgacgccct	ggcggcgcgc	gagaatgaac	900
gccaaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcgga	tgatcgcggc	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcatg	aaatcctggc	cggtttgtct	gatgccaaag	tgccggcctg	1080
gccggccagc	ttggccgctg	aagaaaccga	gcgcgcgcgt	ctaaaaaggt	gatgtgtatt	1140
tgagtataaac	agcttgcgct	atgcggtcgc	tgcttatatg	atgcgatgag	taataataaca	1200
aatacgcgaag	gggaacgcat	gaagggttatc	gctgtactta	accagaaaag	cgggtcaggc	1260
aagacgacca	tcgcaaccga	tctagcccgc	gccctgcaac	tcgccggggc	cgatgttctg	1320
ttagtgcatt	ccgatcccc	gggcagtgcc	cgcgattggg	cggccgtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccgcgcc	acgattgacc	gcgacgtgaa	ggccatcgcc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	ccccaggcgc	cggacttggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatgggccc	1560
accgcccagc	tcgtggagct	gggttaagcag	gcgattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	gggtgcccag	1680
gcgctggccg	ggtacgagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgcgc	cacaaccgtt	cttgaatcag	aaccgcaggg	cgacgctgcc	1800
cgcgagggtcc	aggcgctggc	cgctgaaatt	aaatcaaaa	tcatttgagt	taatgaggta	1860
aagagaaaat	gagcaaaaagc	acaaacacgc	taagtgcggg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgcgcgc	ggagcgtgac	agatgtacgc	agctaccaga	ggtagcccaa	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgccg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagtagat	gaatttttagc	ggctaaagga	ggcgcatggg	aaaatcaaga	acaaccaggc	2160
accgacgcgc	tgggaatgccc	catgtgtgga	ggaacgggcg	gttggccagg	cgtaagcgcc	2220
tgggttgtct	gccggccctg	caatggcact	ggaaccccca	agcccgagga	atcggtgcta	2280
cggctcgaaa	ccatccggcc	cgggtacaaat	cggcgccggc	ctgggtgatg	acctgggtgga	2340
gaagttgaag	gccgcgcagg	ccgcccagcg	gcaacgcac	gaggcagaag	cacgcccccg	2400

tgaatcgtgg	caagcggccg	ctgatogaat	ccgcaaagaa	tcccggcaac	cgccggcagc	2460
cggtgcgccg	tgcattagga	agccgcccac	ggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tgcgagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggt	tccgcagggc	cggccggcat	ggccagtgtg	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgagggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	ggcgggcccg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccggtttc	tctaccgcct	3120
ggcacgccgc	agcgagaagc	agggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	cgggagtacg	atgtgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tccgagggcga	agcatccgcc	ggttccctaat	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccctg	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tcgcctaaa	actcttttaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgtgcgcg	tccctacgcc	ccgcgcgttc	gcctcgccct	atcgcgcccg	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	ggcgccccc	atcaaggcac	cctgcctcgc	gcgtttcggg	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccgggag	cggtcacagc	ttgtctgtaa	gcggatgccc	3900
ggagcagaca	agcccgtcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagccca	3960
tgaccagctc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcccgtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcctc	aggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggtcggttcg	4140
gctcgccgca	cgggtatcag	ctcactcaaa	cgggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtt	ctggcggttt	tccataggct	ccgccccctc	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataaccagg	cgtttccccc	4380
tggaagctcc	ctcgtgcgct	ctcgtgttcc	gacccctgccg	cttaccggat	acctgtccgc	4440
ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgcttc	agcccgaccg	4560
gtcgccctta	tccggttaact	atcgtcttga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggcccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgtcgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacacaa	4800
caccgctggt	agcgggtggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaaactc	4920
acgttaaggg	attttgggtca	tgcatctctg	gtactaaaa	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatatcct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgcctcg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccagg	cgccgtggga	aaagacaaag	tccctcttcg	gcttttccgt	5220
gcaatccaca	tccgcccagat	cgcgcggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
taagctattc	gtatagggac	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
cgcatacagc	tcgataatct	aatccgatct	gtcgatggag	tgaaagagcc	tgatgcactc	5400
gacgccatcg	gcctcactca	tttcagggct	ttgtcatct	tcatactctt	ccgagcaaa	5460
gacctttgga	acaggcagct	tgagcagatt	gctccagcca	tcattgcccgt	caaagtgcag	5520
atcataagtg	gtccctttat	ttccttccag	ccatagcatc	atgtcctttt	cccgtttccac	5580
tcccaccagc	ttatatacct	accggctgtc	cgctattttt	aaatataggt	tttcattttc	5640
tttttccgatc	agttttttca	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tcctcttttc	tacagtattt	attccgggtga	tattctcatt	ttagccattt	attattttct	5760
aattcactgt	tccttgcat	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
ttttcaaagt	tgccgtataa	ctaaaaacct	aaataccaga	aaacagcttt	ttcaaaagttg	5880
caggcgcgca	cgctctgtca	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
gtttcaaac	cgccagctta	tcgttacaa	caacatgcta	ccctccgcga	gatcatccgt	6000
tctgcgcct	tacaacggct	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6060
cgagtgtgga	ttttgtgccg	ctcccgcgtga	cgccgtcccg	gactgatggg	ctgcctgtat	6120
tatatgtgtg	tgtaaacaaa	agctgccggg	cggggagctg	ttgggtggct	gggtggcagga	6180
taattgtactg	aatcaacgcc	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
gttttaggaa	ttagaaat	tattgataga	gggggatctg	gatttttagta	ctggattttg	6300
ggtttcttat	atgctcaaca	catgagcgaa	agttttttac	aaatacaaat	acatactaag	6360
			acctatagg	aacctaat	cccttatctg	6420

ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcgggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtcc	agctggccaga	aaccacgctc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccgt	ccgctgggtg	6720
cgggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccctccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgcga	gacggacgag	gtcgtccgtc	cactccctgog	gttccctgcgg	ctcgggtacgg	6900
aagttgaccg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgacagccgc	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgtcgtttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttata	agaggaaggt	cttgcgaaag	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	gcaagacgtg	gttggaaagt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcttttat	gcgaatgatg	gcattttag	gtgccacctt	ccttttctac	7320
tgtccttttg	atgaagtgc	agatagctgg	ccaatggaa	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaagtct	caatagccct	ttgggtcttct	gagactgtat	ctttgatatt	7440
cttgagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcctgaac	gatagccttt	cctttatcgc	aatgatggca	7620
ttttaggtg	ccaccttcct	tttctactgt	ccttttgatg	aagtgcagca	tagctgggca	7680
atggaaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtattct	tgtatttctt	ggagttagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaagcggg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttagggaccc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	ttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cggatatacag	acatgataag	atacatgat	8100
gagtttggac	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaaatttgt	8160
gatgctattg	ctttatttgt	aaccattata	agctgcaata	aacaagttag	ggtggggcga	8220
gaactccagc	atgagatccc	cgcgctggag	gatcatccag	ccggcgtccc	ggaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgt	gcacgtgtca	8340
gtcctgctcc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	cgctgctcgc	cgactcggg	cacgagcggc	ccggaggcgt	cccggaagtt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctgggtc	ctggaccgcg	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acacggcgca	agtcgtcctc	cagcaagtcc	cgggagaacc	cgagccgggtc	8640
ggtccagaac	tcgaccgctc	cggcgacgtc	gcgcgcgggtg	agcaccgga	cggcactgggt	8700
caacttggcc	atggatccag	atttgcgtca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	gattcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaagggc	tattgagact	tttcaacaaa	gggttaatatc	gggaaacctc	ctcggatttc	8880
attgcccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaaggaaggt	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgccctctgcc	gacagtgggtc	9000
ccaaagatgg	acggagagca	tgtgataaca	tcgttgaaaa	agaagacgtt	ccaaccaggt	9060
cttcaaagca	agtggattga	tcagaagacc	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	ggattccatt	aaagggctat	tgagactttt	caacaaaggg	9180
taatatcggg	aaacctcctc	acctacaaat	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaaggtggc	agtggtccca	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgcccgc	accacgtctt	aagatggacc	cccacccacg	aggagcatcg	9360
tgaaaaaaga	agacgttcca	caatcccact	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaa	ggatgacgca	ggacacgctg	atccttcgca	agaccttcct	ctatataagg	9480
aagttcattt	catctggaga	ccggggggggc	aaatcaccag	tctctctcta	caaattctatc	9540
tctctcgagc	tttcgcagat	ctgatcgaaa	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	gctgctttca	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	gatggtttct	gcttcgatgt	aggaggcgt	ggatatgtcc	9720
tgccggtaaa	tagctgcgcc	ccggaagtgc	acaaagatcg	ttatgtttat	cggcactttg	9780
catcggccgc	gctcccgtat	gcacagggtg	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgt	gtcgcggagg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgccccgtgt	tctacaaccg	ccattcggac	ctatggatgc	gatcgctgog	gccgatctta	9960
gccagacgag	cgggttcggc	gctgatcccc	cgcaaggaa	cgggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gtgcacgcgg	atgtgtatca	ctggcaaaact	gtgatggacg	10080
acaccgtcag	tgctccgtc	gtgacgcgg	atcttcgctc	gatgcttttg	gccgaggact	10140
gccccgaagt	ccggcacctc	attgactgga	gcgaggcgat	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcgggtc	tggaggcgtg	ggttggttg	gttcggggat	tcccaatacg	10260
aggtcgccaa	catcttcttc	gagcttgcag	gatcgccaacg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	tatcagagct	tggttgacgg	actccgggog	tatatgctcc	10380
gcattggtct	tgaccaactc			caatttcgat	gatgcagctt	10440

ggggcgagg	tcgatgcgac	gcaatcgctc	gatccggagc	cgggactgtc	gggcgtacac	10500
aaatcgccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaacc	acgccccagc	actcgctcga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagttttctc	ataataatgt	gtgagtagtt	cccgataaag	10680
ggaatttagg	ttccctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaattttc	aatttcctaaa	acccaaatcc	10800
agtactaaaa	tccagatccc	cogaattaat	tccggcttaa	ttcagatcaa	gcttggcact	10860
ggccgctcgt	ttcaaacgct	gtgactggga	aaaccctggc	gttaccacac	ttaatcgcc	10920
tgcagcacat	ccccctttcg	ccagctggcg	taataagcga	gaggcccgca	ccgatcgccc	10980
ttcccaacag	ttgcgcgagc	tgaatggcga	atgctagagc	agcttgagct	tggatcagat	11040
tgctggtttc	cgccttcagt	ttaaactatc	agtgtttgac	aggatatatt	ggcgggtaaa	11100
ctaagagaaa	aagagcggtt	attagaataa	cggatattta	aaagggcgctg	aaaaggttta	11160
tccattctatc	catcttatatg	tg				11182

```
<210> 90
<211> 8428
<212> DNA
<213> Artificial Sequence
```

<220>
<223> pCambia3300 Plasmid

[illegible]

cgtagaggtt	tccgcagggc	cgcccgccat	ggccagtgtg	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cggaagggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgtc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cgccgccttg	gtgacggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	ggcgggcccg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggtacgtgct	3060
gacggttcac	cccgtattact	ttttgatcga	tcccggcatc	ggcgttttct	tctaccgcct	3120
ggcacgcgcg	gcccgcaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtacg	atgtgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgttac	cgcaacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattcgggaac	cggaacccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	actgtaagtg	actgataata	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actcttataa	catctataaa	actctataaa	cccgctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgcgttc	gcgtcggcct	atcgcgggcg	ctggccgctc	3720
aaaaatggct	ggcgctcagg	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	ggcgcccccac	atcaaggcac	cctgcctcgc	gcgtttcggt	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccgga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cggtgtgtgg	cggtgtgtcg	ggcgagccca	3960
tgaccacagt	agcggagtgt	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcata	aggcgctctt	ccgcttcctc	gctcaactgac	tcgctgcgct	cggtcggttcg	4140
gctgcggcga	gcggtatcac	ctcactcaaa	ggcggtataa	cggttatcca	cagaatcagg	4200
ggataaacga	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcggttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttcccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgtccg	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccggttc	agccccgaccg	4560
ctgcgcctta	tccggttaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tgggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgtgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaaaaaac	4800
caccgctggg	agcgggtgggt	tttttgggtt	caagcagcag	attacgcgca	gaaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	gggtctgac	gctcagtgga	acgaaaactc	4920
acgttaaggg	atttttggta	tgcattctag	gtactaaaaa	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgtttcttc	ccgatatcct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgcccctg	ccgcttctcc	caagatcaat	aaagccactt	actttggccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgcggtggga	aaagacaagt	tcctcttcgg	gcttttcctg	5220
ctttaaaaaa	tacataagct	cgcgcggatc	tttaaatgga	gtgtcttctt	ccaggttttc	5280
gcaatccaca	tcggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctatcc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcatct	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcataggtg	gtccccttat	accggctgtc	cgctattttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggtg	5700
tttttcgata	agttttttca	attccgggtg	tattctcatt	ttagccattt	attatttctc	5760
tctctctttc	tcagtatttt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tcttgcattt	ctaaaacctt	aaataaccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgcggttc	ttccgaatag	catcggtaac	atgagcaaa	6060
tctgcgcgct	tacaacggct	ctcccgtga	cgcggtcccg	gactgatggg	ctgcctgtat	6120
cgagtgggtg	ttttgtgccc	agctgcgggt	cggggagctg	ttggctgggt	ggtggcagga	6180
tatatgtgtg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggtcgtttt	6240
taattgtactg	aattaacgcc	gaattaatc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acataactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	acctatagg	aaccctaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcagatcaaa	tctcggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagttc	agctgccaga	aaccacgctc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tcgggtcgtt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660

gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcaccgt	ccgctgggtgg	6720
cgggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccttccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgca	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgctg	ctcgggtacgg	6900
aagttgaccg	tgcttgctc	gatgtagtgg	ttgacgatgg	tgacagaccg	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgctcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatata	agaggaaggt	cttgcggaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcccttat	cgcaatgatg	gcatttgtag	gtgccacctt	ccttttctac	7320
tgtccttttg	atgaagtgc	agatagctgg	gcaatggaat	ccgaggagg	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggctctct	gagactgtat	ctttgatatt	7440
cttgagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
ttttaggtg	ccaccttctt	tttctactgt	ccttttgatg	aagtgcaga	tagctgggca	7680
atggaatccg	aggagggttt	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgtcttagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaaagcgg	gcagtgcagc	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tccggctcgt	7980
atgttggtg	gaattgtgag	cggataacaa	tttcaacacg	gaaacagcta	tgaccatgat	8040
tacgaattcg	agctcggtag	ccggggatcc	tctagagtcg	acctgcaggc	atgcaagctt	8100
ggcactggcc	gtcgtttttac	aacgtcgtga	ctgggaaaaac	cctggcgtta	cccaacttaa	8160
tcgccttgca	gcacatcccc	ctttcgcag	tggcgtaaat	agcgaagagg	ccgcaccga	8220
tcgccccttc	caacagttgc	gcagcctgaa	tggcgcaatgc	tagagcagct	tgagcttgga	8280
tcagattgtc	gtttcccgcc	ttcagtttaa	actatcagtg	tttgacagga	tatatggcg	8340
ggtaaacccta	agagaaaaga	gcgtttatta	gaataacgga	tatttaaaa	ggcgtgaaaa	8400
ggtttatccg	ttcgtccatt	tgtatgtg				8428

<210> 91

<211> 3438

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBZeo Plasmid

<400> 91

tcgaccctct	agtcaaggcc	ttaagtgagt	cgtattacgg	actggccgctc	gtttttacaac	60
gtcgtgactg	ggaaaaacct	ggcgttaccc	aacttaaatcg	ccttgacgca	catccccctt	120
tcgccagctg	gcgtaataagc	gaagaggccc	gcaccgatcg	cccttcccaa	cagttgcgca	180
gcctgaatgg	cgaatggcgc	ttcgccttgg	aataaaagccc	gcttcggcgg	gctttttttt	240
gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgttttatt	300
ttttctaaata	cattcaaata	tgtatccgct	catgagacaa	taacctgat	aatgcttca	360
ataaatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	420
ttttgcgcca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctggtga	aagtaaaaga	480
tgtgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	540
gatccttgag	agttttcgcc	cogaagaacg	ttctccaatg	atgagcactt	ttaaagttct	600
gctatgtggc	gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgccgcat	660
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	720
tggcatgaca	gtaagagaat	tatgcagtcg	tgccataacc	atgagtgata	acactgcggc	780
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	840
gggggcatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	900
cgacgagcgt	gacaccacga	tgctgtagc	aatggcaaca	acgttgcgca	aactattaac	960
tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	1020
agttgcagga	ccactctctg	gctcgccctt	tccggctggc	tggtttattg	ctgataaatc	1080
tggagccggg	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	1140
ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	1200
acagatcgct	gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	1260
ctcatatata	ctttagattg	atttaccctg	gttgataatc	agaaaagccc	caaaaacagg	1320
aagattgtat	aagcaaatat	ttaaatgtga	aacgttaata	ttttgttaaa	attcgcgtta	1380
aatttttgtt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1440
aaatcaaaaag	aatagcccga	gataggggtg	agtgttggtc	cagtttgga	caagagtcca	1500
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1560
ccactacgtg	aaccatcacc	caaatacaagt	tttttggggt	cgaggtgccg	taaagcacta	1620

aatcgggaacc	ctaaagggag	ccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1680
gaaagggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcggcca	1740
cgctgcgcgt	aaccaccaca	cccgcgcgcg	ttaatgcgcc	gctacagggc	gcgtaaaagg	1800
atctaggtga	agatcctttt	tgataatctc	atgaccaaaa	tcccttaacg	tgagttttcg	1860
ttccactgag	cgtcagaccc	cgtagaaaag	atcaaaggat	cttcttgaga	tccttttttt	1920
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttggttg	1980
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	2040
ccaaatactg	ttctttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	2100
ccgcctacat	acctoegtct	gctaattcctg	ttaccagtgg	ctgctgccag	tggcgataag	2160
tcgtgtctta	ccgggttggg	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	2220
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2280
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aaggagagaaa	ggcggacagg	2340
tatccggtaa	gcggcagggt	cggaacagga	gagcgcacga	gggagcttcc	agggggaaac	2400
gcctgggtatc	tttatagctc	tgtcgggttt	cgccacctct	gacttgagcg	tcgatttttg	2460
ttgtgctcgt	cagggggggc	gagcctatgg	aaaaacgcc	gcaacgcggc	ctttttacgg	2520
ttcctggcct	tttgctggcc	tttgctcac	atgtaatgtg	agttagctca	ctcattaggg	2580
acccccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcggata	2640
acaattttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atagcactca	2700
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gattgaagcc	2760
tgcttttttta	tactaacttg	agcgaaatct	ggatccatgg	ccaagttgac	cagtgccgtt	2820
ccgggtgctca	ccgcgcgcga	cgtcgccgga	gcggctcgagt	tctggaccga	ccggctcggg	2880
ttctcccggg	acttcgtgga	ggacgacttc	gccgggtgtg	tccgggacga	cgtgaccctg	2940
ttcatcagcg	cgggtccagg	ccagggtgtg	ccggacaaca	ccctggcctg	gggtgtgggtg	3000
cgcgccctgg	acgagctgta	cgccgagtg	tcggaggtcg	tgtccacgaa	cttccgggac	3060
gcctccgggc	cggccatgac	cgagatcggc	gagcagccgt	gggggcgggg	gttcgccttg	3120
cgcgaccggc	ccggcaactg	cgtgcacttc	gtggccgagg	agcaggactg	acacgtgcta	3180
cgagatttcg	attccaccgc	cgcttctat	gaaagggttg	gcttcggaat	cgttttccgg	3240
gacgccggct	ggatgatcct	ccagcgcggg	gatctcatgc	tggagttctt	cgcccacccc	3300
aacttgttta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	3360
aataaagcat	ttttttcact	gcattctagt	tgtggtttgt	ccaaactcat	caatgtatct	3420
tatcatgtct	gtataccg					3438

<210> 92
 <211> 10549
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pCambia1302 Plasmid

<300>
 <308> Genbank #AF234398
 <309> 2000-04-24

<400> 92						
catggttagat	ctgactagta	aaggagaaga	acttttctact	ggagttgtcc	caattcttgt	60
tgaattagat	ggtgatgtta	atgggcacaa	attttctgtc	agtggagagg	gtgaagggtga	120
tgcaacatac	ggaaaaactta	cccttaaatt	tatttgcact	actggaaaac	tacctgttcc	180
gtggccaaca	cttgtaacta	ctttctctta	tggtgttcaa	tgcttttcaa	gatacccaga	240
tcatatgaag	cggcagcact	tcttcaagag	cgccatgcct	gagggatacg	tgaggagag	300
gaccatcttc	ttcaaggacg	acgggaacta	caagacacgt	gctgaagtca	agtttgaggg	360
agacaccctc	gtcaacagga	tcgagcttaa	gggaatcgat	ttcaaggagg	acggaaacat	420
cctcggccac	aagttggaat	acaactacaa	ctcccacaac	gtatacatca	tggccgacaa	480
gcaaaagaac	ggcatcaaag	ccaacttcaa	gaccggccac	aacatcgaag	acggcggcgt	540
gcaactcgct	gatcattatc	aacaaaatac	tccaattggc	gatggccctg	tccttttacc	600
agacaacccat	tacctgtcca	cacaatctgc	cctttcgaaa	gatcccaacg	aaaagagaga	660
ccacatggtc	cttcttgagt	ttgtaacagc	tgctgggatt	acacatggca	tgatgaact	720
atacaaagct	agccaccacc	accaccacca	cgtgtgaatt	ggtgaccagc	tcgaatttcc	780
ccgatcgctc	aaacatttgg	caataaagtt	tcttaagatt	gaatcctgtt	gccggctctg	840
cgatgattat	catataattt	ctggtgaatt	acgttaagca	tgtataaatt	aacatgtaatt	900
gcatgacggt	atztatgaga	tgggttttta	tgattagagt	cccgaatta	tacatttaatt	960
acgcgataga	aaacaaaata	tagcgcgcaa	actaggataa	attatcgcg	gcgggtgtcat	1020
ctatgttact	agatcgggaa	ttaaactatc	agtgtttgac	aggatatatt	ggcgggttaa	1080
cctaagagaa	aagagcggtt	attagaataa	cggatattta	aaagggcggt	aaaaggttta	1140
tcggttcgtc	catttgtagt	tgcatgccaa	ccacagggtt	cccctcgggg	tcaaagtaact	1200
ttgatccaac	ccctccgctg	ctatagtgc	gtcggcttct	gacgttcagt	gcagccgctc	1260
tctgaaaacg	acatgtcgca	caagtcctaa	gttacgcgac	aggctgcccc	cctgcccttt	1320

tcctggcggtt	ttcttgtcgc	gtgttttagt	cgcataaagt	agaatacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgcccgcgtc	1440
agcaccgacg	accaggactt	gaccaaccaa	cgggcccgaac	tgacgcggcg	cggctgcacc	1500
aagctgtttt	ccgagaagat	caccgggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgcccctg	cgacgttgtg	acagtgacca	ggctagaccg	cctggcccgc	1620
agcaccgcg	acctactgga	cattgcccag	cgcacccagg	aggccggcgc	gggcccgtgc	1680
agcctggcag	agccgtgggc	cgacaccacc	acgcgggccc	gccgcctggt	gttgaccgtg	1740
ttcgccggca	ttgcccagtt	cgagcggtcc	ctaactcatc	accgcaccgc	gagcggggcg	1800
gaggccgcca	aggcccagag	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgcgcacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggctgca	1920
ctgcttggcg	tgcatcgctc	gaccctgtac	cgcgcacttg	agcgcagcga	ggaagtgcag	1980
cccaccggag	ccaggcgggc	cgggtgcctt	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgccgc	ccgagaatga	acgccaaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcgaggcgga	gatgatcgcg	gccgggtacg	2160
tgctcgagcc	gcccgcgcac	gtctcaaccg	tgccggtgca	tgaaatcctg	gccgggttgt	2220
ctgatgcaa	gctggcgccc	tgcccgggcc	gcttggccgc	tgaagaaacc	gagcggcgcc	2280
gtctaaaaag	gtgatgtgta	tttgagtaaa	acagcttgcg	tcctgcccgc	gctgctgata	2340
tgatgcgatg	agtaataaaa	caaatacgca	aggggaacgc	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcgcccgtga	2460
actcgccggg	ccgcatgttc	tggttagtcga	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcggccggt	cggaagatc	aaccgctaac	cgttgtcgcc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gccggcgccg	cttcgtagtg	atcgaccggg	cgccccaggc	2640
ggcggacttg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cgggtgcagc	2700
aagcccttac	gacatatggg	ccaccgcgga	cctgggtggg	ctgggttaagc	agcgcattga	2760
ggctacggat	ggaaggctac	aagcggcctt	tgctgtgtcg	cgggcgatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	aggcgctggc	ggggtacgag	ctgcccattc	ttgagtcagg	2880
tatcacgcag	cgcgtgagct	acccaggcac	tgccgcggcc	ggcacaaccg	ttcttgaatc	2940
agaacccgag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtggc	3060
ggccgtccga	gcgcacgcag	cagcaaggct	gcaacgcttg	ccagcctggc	agacacgcca	3120
gccatgaagc	gggtcaactt	tcagttgccc	cgaggagatc	acaccaagct	gaagatgtac	3180
gcggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aatgagtag	atgaatttta	gcccgtaaaag	gaggcgccat	3300
ggaaaaatcaa	gaacaaccag	gcaccgacgc	cttggaatgc	cccatgtgtg	gaggaaacggg	3360
cgggttgcca	ggcgtaagcg	gctgggttgt	ctgcccggcc	tgcaatggca	ctggaacccc	3420
caagcccagag	gaatcgccgt	gacgggtcgca	aaccatccgg	cccgggtacaa	atcgccgcgc	3480
cgtgggtgga	tgacctgggt	gagaagtga	aggccgcgca	ggccgcccag	cggcaacgca	3540
tcgaggcaga	agcacgcgcc	ggtagaatcgt	ggcaagcggc	cgctgatcga	atccgcaaaag	3600
aatcccggca	accgcgggca	gccgggtgcgc	cgtcgattag	gaagccgccc	aaggggcgacg	3660
agcaaccaga	ttttttcggt	ccgatgctct	atgacgtggg	caccgcgat	agtcgcagca	3720
tcattggacgt	ggccgttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tccagacggg	cacgtagagg	tttccgcagg	gccggccggc	atggccagtg	3840
tgtgggatta	cgacctggta	ctgatggcg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaaggagagc	aagcccggcc	cgctgttccg	tcacacgctt	gcccagctac	3960
tcaagttctg	ccggcgagcc	gatggcgga	agcagaaaaga	cgacctggta	gaaacctgca	4020
ttcggttaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggt	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcggcc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacgggtc	accccgatta	ctttttgatc	gatcccggca	4260
tcggccgttt	tctctaccgc	ctggcacgcc	gcgcccagag	caaggcagaa	gccagatggt	4320
tgttcaagac	gatctacgaa	cgcagtgga	gcgcccggga	gttcaagaag	ttctgtttca	4380
ccgtgcccga	gctgatcggg	tcaaatgacc	tgcccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatg	cctagcaggg	gaaaaagggtc	4560
gaaaaggctct	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtacattgg	gaaccggtca	cacatgtaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgcta	aaactcttta	aaacttatta	4740
aaactcttaa	aaccgcctg	gcctgtgcat	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggctcgctgc	gctccctacg	ccccgcgctc	tcgctcgggc	4860
ctatcgcggc	cgtgtggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcccgaacgc	cgccgcgtcg	ccactcgacc	gcccgcggcc	acatcaaggc	accctgcctc	4980
gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacggtcaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgctc	agggcgcgctc	agcgggtggt	5100
ggcgggtgtc	ggggcgcgagc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgc	ggcatcagag	cagattgtac	tgagagtgc	ccatagcgcg	tgtgaaatac	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggctgcggc	gagcgggtatc	agctcactca	aaggcggttaa	5340

tacgggttato	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	gcgaaaccgg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgtagc	ctctcctgtt	ccgaccctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aacccccgt	tcagcccgac	cgctgcgctc	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggtaaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggtatgtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggcaaaacaa	accaccgctg	gtagcggtagg	tttttttgtt	tgcaagcagc	6000
agattacgcy	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acggggtctg	6060
acgctcagtg	gaacgaaaaac	tcacggttaag	ggatttttgg	catgcattct	aggtaactaa	6120
acaattcatc	cagtaaaata	taatatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatata	ctccctgatc	gaccggagcg	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgcc	atctttcaca	aagatgttgc	tgctctccag	gtcgcctggg	gaaaagacaa	6360
gttccctctc	gggctttttcc	gtcttttaaaa	aatcatacag	ctcgcgcgga	tcttttaaag	6420
gagtgctctc	ttcccagttt	tcgcaatcca	catcggccag	atcggtattc	agtaagtaac	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tcgcataaca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccagagcaa	aggacgccat	cgccctcact	catgagcaga	ttgctccagc	6660
catcatgccc	ttccaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcagtctctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tcgctcattt	6780
ttaaatatag	gttttcatct	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tattttttcga	tcagtttttt	caattccggg	gatattctca	6900
ttttagccat	ttattatttc	cttctctttt	tctacagtat	ttaaagatag	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttcccttgc	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tggttttcaaa	gttggcggtat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	atctgctctgt	catcggttaca	atcaacatgc	7140
taccctccgc	gagatcatcc	gtgtttcaaaa	cccggcagct	tagttgcccgt	tcttccgaat	7200
agcatcggta	acatgagcaa	agtctgcccgc	cttacaacgg	ctctcccgc	gacgcgcgtc	7260
cggactgatg	ggctgcccgt	atcgagtggt	gattttgtgc	cgagctgccc	gtcggggagc	7320
tggtggctgg	atgtgtggcag	gatataattgt	gggttaaaaca	aattgacgct	tagacaaact	7380
aataacacat	tgccgacgct	tttaattgtac	tgaattaaag	ccgaattaat	tcgggggatc	7440
tggatttttag	tactggattt	tggtttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	aatgctcaa	cacatgagcg	aaacctata	7560
ggaacctaa	ttcccttatc	tggaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccggct	ggcatctact	ctatttcttt	gcccctcgac	gagtgctggg	7680
gcgtcgggtt	ccactatcgg	cgagtacttc	tacacagcca	tcgggtccga	cgcccgcgct	7740
tttgccggcg	atgtgtgtac	gcccagacag	cccggctccg	gatcggagca	ttgctgcgca	7800
tcgaccctgc	gcccagagct	catcatcgaa	attgccgtca	accaagctct	gatagagttg	7860
gtcaagacca	atgcccagca	tatacgccc	gagtcgtggc	gatcctgcaa	gctccggatg	7920
cctccgctcg	aagtagcgcg	tctgctgctc	catacaagcc	aaccacggcc	tccagaagaa	7980
gatgttggcg	acctcgtatt	gggaatcccc	gaacatcgcc	tcgctccagt	caatgaccgc	8040
tggtatgccc	ccattgtccg	tcaggacatt	gttggagccc	aaatccgcgt	gcacgaggtg	8100
ccggacttcg	gggcagtcct	cggcccaaaag	catcagctca	tcgagagcct	gcgcgacgga	8160
cgcactgacg	gtgtcgtcca	tcacagtttg	ccagtgtata	acatggggat	cagcaatcgc	8220
gcatatgaaa	tcacgcccag	tagtgtattg	accgattcct	tgccgtccga	atgggcccga	8280
cccgtcctgc	tggttaagat	cggccgcagc	gatcgcaccc	atagcctccg	cgaccgggtg	8340
tagaacagcg	ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccct	gtgcacggcg	8400
ggagatgcaa	taggtcaggc	tctcgctaaa	ctccccaatg	tcaagcactt	ccggaatcgg	8460
gagcgcggcc	gatgcaaaat	gcccagataaa	ataacgatct	ttgtagaaac	catcggcgca	8520
gctattttacc	cgcaggacat	atccacgccc	tcctacatcg	aagctgaaag	cacgagattc	8580
ttcgcctctc	gagagctgca	tcagggtcgga	gacgtgtctg	aacttttctg	tcagaaactt	8640
ctcgacagac	gtcgcgggtga	gttcaggctt	tttcatatct	cattgcccc	cgggatctgc	8700
gaaagctcga	gagagataga	ttttagagaga	gagactgggt	atttcagcgt	gtcctctcca	8760
aatgaaatga	acttctttat	atagaggaag	gtcttgcgaa	ggatagtggt	attgtgcgtc	8820
atcccttacg	tcagtggaga	tatcacatca	atccacttgc	tttgaagacg	tggttggaac	8880
gtcttctttt	tcacagatgc	tcctcgtagg	tggggggtcca	tctttgggac	cactgtcggc	8940
agaggcatct	tgaacgatag	cctttctctt	atcgcaatga	tggcatttgt	aggtgccacc	9000
ttccttttct	actgtccttt	tgatgaaagt	acagatagct	gggcaatgga	atccgaggag	9060
gtttcccgat	attacccttt	gttgaaaagt	ctcaatagcc	ctttggtctt	ctgagactgt	9120
atctttgata	ttcttgaggt	agacgagagt	gtcgtgctcc	accatgttat	cacatcaatc	9180
cacttgcttt	tggaacgtgg	ttggaaacgtc	ttcttttctc	acgatgctcc	tcgtgggtgg	9240
gggtccatct	ttgggaccac	tgtagccagca	ggcatcttga	acgatagcct	ttcctttatc	9300
gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	tgaagtgaca	9360

gatagctggg	caatggaatc	cgaggaggtt	tcccgatatt	accctttgtt	gaaaagtctc	9420
aatagccctt	tgggtcttctg	agactgtatc	tttgatattc	ttggagtaga	cgagagtgtc	9480
gtgctccacc	atgtttggcaa	gctgctctag	ccaatacgca	aaccgcctct	ccccgcgcgt	9540
tggccgattc	attaatgcag	ctggcacgac	aggtttcccg	actggaaagc	gggcagttag	9600
cgcaacgcaa	ttaatgtgag	ttagctcact	cattagggac	cccaggcttt	acactttatg	9660
cttccggctc	gtatgtttgt	tggaaattgtg	agcggataac	aatttcacac	aggaaacagc	9720
tatgaccatg	attacgaatt	cgagctcggt	acccggggat	cctctagagt	cgacctgcag	9780
gcatgcaagc	ttggcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accctggcgt	9840
taccaacttt	aatcgccctt	cagcacatcc	ccctttcgcc	agctggcgta	atagcgaaga	9900
ggcccgccacc	gatcgccctt	cccaacagtt	gcccagcctg	aatggcgaat	gctagagcag	9960
cttgagcttg	gatcagattg	tcgtttcccg	ccttcagttt	agcttcatgg	agtcaaagat	10020
tcaaataagag	gacctaacag	aactcgccgt	aaagactggc	gaacagttca	tacagagtct	10080
cttacgactc	aatgacaaga	agaaaatctt	cgtcaacatg	gtggagcacg	acacacttgt	10140
ctactccaaa	aatatcaaag	atacagtcct	agaagaccaa	agggcaattg	agacttttca	10200
acaaagggtg	atatccggaa	acctcctcgg	attccattgc	ccagctatct	gtcactttat	10260
tgtgaagata	gtggaaaagg	aaggtggctc	ctacaaatgc	catcattgcg	ataaaaggaaa	10320
ggccatcggt	gaagatgcct	ctgccgacag	tgggtcccaa	gatggacccc	caccacagag	10380
gagcatcggt	gaaaaagaag	acgttccaac	cacgtcttca	aagcaagtgg	attgatgtga	10440
tatctccact	gacgtaaggg	atgacgcaca	atcccactat	ccttcgcaag	acccttcctc	10500
tatataagga	agttcatttc	atttggagag	aacacggggg	actcttgac		10549

<210> 93
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> CaMV35SpolyA Primer

<400> 93		
ctgaattaac	gccgaattaa	ttcgggggat ctg
		33

<210> 94
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> CaMV35Spr Primer

<400> 94		
ctagagcagc	ttgccaacat	ggtggagca
		29

<210> 95
 <211> 12592
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pAg2 Plasmid

<400> 95						
gtacgaagaa	ggccaagaac	ggccgcctgg	tgacgggtatc	cgagggtgaa	gccttgatta	60
gccgctacaa	gatcgtaaa	agcgaaccgg	ggcggccgga	gtacatcgag	atcgagctag	120
ctgattggat	gtaccgcgag	atcacagaag	gcaagaaccc	ggacgtgctg	acggttcacc	180
ccgattactt	tttgatcgat	cccggcatcg	gccgttttct	ctaccgcctg	gcacgccgcg	240
ccgcaggcaa	ggcagaagcc	agatggttgt	tcaagacgat	ctacgaacgc	agtggcagcg	300
ccggagagtt	caagaagttc	tgtttcaccc	tgccgaagct	gatcgggtca	aatgacctgc	360
cggagtacga	tttgaaggag	gaggcggggc	aggctggccc	gatcctagtc	atgcgctacc	420
gcaacctgat	cgaggggcga	gcacccgcgg	gttcctaatt	tacggagcag	atgctagggc	480
aaattgccct	agcaggggaa	aaaggtcgaa	aaggtctctt	tcctgtggat	agcacgtaca	540
ttgggaaccc	aaagccgtac	attgggaacc	ggaaccgcta	cattgggaac	ccaaagccgt	600
acattgggaa	ccggtcacac	atgtaagtga	ctgatataaa	agagaaaaaa	ggcgattttt	660
ccgcctaaaa	ctctttaaaa	cttattaaaa	ctcttaaaac	ccgcctggcc	tgtgcataac	720
tgtctggcca	gcgcacagcc	gaagagctgc	aaaaagcgcc	tacccttcgg	tcgctgcgct	780
ccctacgccc	cgccgcttcg	cgtcggccta	tcgcccgcgc	tggccgctca	aaaatggctg	840
gcctacggcc	aggcaatcta	ccagggcgcg	gacaagccgc	gccgtcgcca	ctcgaccgcc	900

ggcgccacaca	tcaaggcacc	ctgcctcgcg	cgtttcgggtg	atgacgggtga	aaacctctga	960
cacatgcagc	tcccgaggagc	ggtcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	1020
gcccgtcagg	gcgcgtcagc	gggtgttggc	gggtgtcggg	gcgcagccat	gacccagtca	1080
cgtagcgata	gcggagtgtg	tactggctta	actatgcggc	atcagagcag	attgtactga	1140
gagtgaccca	tatgcggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	1200
ggcgctcttc	cgcttccctcg	ctcactgact	cgctgcgctc	ggtcggtcgg	ctgccccgag	1260
cggatatcagc	tcactcaaag	gcggtaatac	ggttatccac	agaatcaggg	gataacgcag	1320
gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgttgc	1380
tggcggttttt	ccataggcto	cgccccctg	acgagcatca	caaaaatcga	cgctcaagtc	1440
agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	1500
togtgcgctc	tcctgttccg	accctgccgc	ttaccggata	cctgtccgcc	tttctccctt	1560
cgggaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	gtgtaggctcg	1620
ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	gcccagaccgc	tgcgccttat	1680
cgggtaacta	tcgtcttgag	tccaacccgg	taagacacga	cttatcgcca	ctggcagcag	1740
ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	ttcttgaagt	1800
ggtaggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	1860
cagttacctt	cggaaaaaga	gttggttagct	cttgatccgg	caaacaaacc	accgctggta	1920
gcgggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaag	1980
atcctttttt	cttttctacg	gggtctgacg	ctcagtgga	cgaaaactca	cgtaaaggga	2040
ttttgggtcat	gcattctagc	tactaaaaa	attcatccag	taaaatataa	tattttattt	2100
totcccaatc	aggcttgatc	cccagtaagt	caaaaaatag	ctcgacatac	tgttcttccc	2160
cgatattcctc	cctgatcgac	cggacgcaga	aggcaatgtc	ataccacttg	tccgccctgc	2220
cgctttctccc	tccttccagc	aagccactta	cttgccatc	tttcacaaag	atgttgctgt	2280
ctcccagggtc	gccgtgggaa	aagacaagtt	cctcttcggg	cttttccgtc	tttaaaaaat	2340
catacagctc	gcgcggatct	ttaaattggag	tgtcttcttc	ccagttttcg	caatccacat	2400
cggccagatc	gttattcagt	aagtaatcca	attcggctaa	gcggctgtct	aagctattcg	2460
tatagggaca	atccgatatg	tcgatggagt	gaaagagcct	gatgcactcc	gcatacagct	2520
cgataatctt	ttcagggtct	tgttcatctt	catactcttc	cgagcaaaag	acgccatcgg	2580
cctcactcat	gagcagattg	ctccagccat	catgccgttc	aaagtgcagg	acctttggaa	2640
caggcagctt	tccttccagc	catagcatca	tgtccttttc	ccgttccaca	ccatagggtg	2700
tccctttata	ccggctgtcc	gtcattttta	aatatagggt	ttcattttct	cccaccagct	2760
tatatacctt	agcaggagac	attccttccg	tatcttttac	gcagcgggat	ttttcgatca	2820
gttttttcaa	ttccggtgat	attctcattt	tagccattta	ttatttctct	cctcttttct	2880
acagtatttta	aagatacccc	aagaagctaa	ttataacaag	acgaactcca	attcactggt	2940
ccttgcatttc	taaaacctta	aataccagaa	aacagctttt	tcaaaagttg	tttcaaagtt	3000
ggcgtataaac	atagtatcga	cggagccgat	tttgaaaccg	cggtgatcac	aggcgaac	3060
gctctgtcat	cgttacaatc	aacatgtac	cctccgcgag	atcatccgtg	tttcaaacc	3120
ggcagcttag	ttgccgttct	tccgaatagc	atcggtaaca	tgagcaaaag	ctgccgcctt	3180
acaaaggctc	tcccgctgac	gccgtcccgg	actgatgggc	tgccgtgtatc	gagtgggtgat	3240
tttgttccga	gctgcgggtc	ggggagctgt	tggctggctg	gtggcaggat	atattgtgggt	3300
gtaaaacaaat	tgacgcttag	acaacttaat	aacacattgc	ggacgttttt	aatgtactga	3360
attaacgccc	aattaattcg	ggggatctgg	atttttagtac	tggatttttg	ttttaggaat	3420
tagaaatttt	attgatagaa	gtattttaca	aatacaataa	cataactaag	gtttcttata	3480
tgctcaacac	atgagcgaaa	ccctatagga	accctaattc	ccttatctgg	gaactactca	3540
cacattatta	tggagaaact	cgagtcaa	ctcggtgacg	ggcaggaccg	gacggggcgg	3600
taccggcagg	ctgaagtcca	gctgccagaa	acccacgtca	tgccagttcc	cgtgcttgaa	3660
gccggccgcc	cgcagcatgc	cgcggggggc	atatccgagc	gcctcgtgca	tgcgcacgct	3720
cgggtcgttg	ggcagcccga	tgacagcgac	cacgctcttg	aagccctgtg	cctccaggga	3780
cttcagcagg	tgggtgtaga	gcgtggagcc	cagtcocgtc	cgctggtggc	ggggggagac	3840
gtacacgggtc	gactcggccg	tccagtcgta	ggcgttgctg	gccttccagg	ggcccgcgta	3900
ggcgatgccg	gcgacctcgc	cgtccacctc	cagggatagc	cagggatagc	gctccgcag	3960
acggacgagg	tcgtccgtcc	actcctgcgg	ttcctgcggc	tcggtacgga	agttgaccgt	4020
gcttgtctcg	atgtagtggg	tgacgatggg	gcagaccgcc	ggcatgtccg	cctcgggtggc	4080
acggcggatg	tcggccgggc	gtcgttctgg	gctcatggta	gactcgagag	agatagattt	4140
gtagagagag	actgggtgatt	tcagcgtgtc	ctctccaaat	gaaatgaact	tccttatata	4200
gaggaaggtc	ttgcgaagga	tagtgggatt	gtgcgtcatc	ccttacgtca	gtggagatata	4260
cacattcaatc	cacttgcttt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	4320
tcgtgggttg	gggtccatct	ttgggaccac	tgtcggcaga	ggcatcttga	acgatagcct	4380
ttcctttatc	gcaatgatgg	catttgtagg	tgccaccttc	ctttctact	gtccttttga	4440
tgaagtga	gatagctggg	caatggaatc	cgaggagggt	tcccgatatt	acccttttgt	4500
gaaaagctctc	aatagccctt	ttgtcttctg	agactgtatc	tttgatattc	ttggagtga	4560
cgagagtgtc	gtgctccacc	atgttatcac	atcaatccac	ttgctttgaa	gacgtgggtg	4620
gaacgtcttc	tttttccacg	atgctcctcg	ttgggtgggg	tccatctttg	ggaccactgt	4680
cggcagagtc	atctgaaacg	atagccttct	ctttatcgca	atgatggcat	ttgtaggtgc	4740
caccttctct	ttctactgtc	cttttgatga	agtgacagat	agctgggcaa	tgggaatccga	4800
ggaggtttcc	cgatattacc	ctttgttgaa	aagtctcaat	agcccttttg	tcttctgaga	4860
ctgtatcttt	gatattcttg	gagtagacga	gagtgtcgtg	ctccaccatg	ttggcaagct	4920

gctctagcca	atacgcaaac	cgccctctccc	cgcgcggttg	ccgattcatt	aatgcagctg	4980
gcacgacagg	tttcccgaact	ggaaagcggg	cagtgcagcg	aacgcaatta	atgtgagtta	5040
gctcactcat	taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tgttggtgg	5100
aattgtgagc	ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgaattcga	5160
gccttgacta	gagggctcgac	ggtatacaga	catgataaga	tacattgatg	agtttggaca	5220
aaccacaact	agaatgcagt	gaaaaaaatg	ctttatgtgt	gaaatttgtg	atgctattgc	5280
tttatttgta	accattataa	gctgcaataa	acaagtggg	gtggggaag	aactccagca	5340
tgagatcccc	gcgctggagg	atcatccagc	cgcgctccc	gaaaacgatt	ccgaagccca	5400
acctttcata	gaaggcggcg	gtggaatcga	aatctcgtag	cacgtgtcag	tcctgtcct	5460
cgccacgaa	gtgcacgcag	ttgcccggcg	ggtcgcgcag	ggcgaactcc	cgccccacg	5520
gctgctcgcc	gatctcggtc	atggccggcc	cggaggcgtc	ccggaagtcc	gtggacacga	5580
cctccgacca	ctcggcgatc	agctcgtcca	ggccgcgcac	ccacaccag	gccagggtgt	5640
tgtccggcac	cacctggtcc	tggaccgcgc	tgatgaacag	gggtcacgtc	ttcccggacca	5700
caccggcgaa	gtcgtcctcc	acgaagtccc	gggagaaccc	gagccggctc	gtccagaact	5760
cgaccgctcc	ggcgacgtcg	cgcgcggtga	gcaccgggaac	ggcactggtc	aacttggcca	5820
tgatccaga	tttcgctcaa	gttagtataa	aaaaagcaggc	ttcaatcctg	cagggaattcg	5880
atcgacactc	tcgtctactc	caagaatatc	aaagatacag	tctcagaaga	ccaaagggct	5940
attgagactt	ttcaacaaaag	ggtaatatcg	ggaaacctcc	tcggattcca	ttgcccagct	6000
atctgtcact	tcattcaaaag	gacagtagaa	aaggaaagggt	gcacctacaa	atgccatcat	6060
tgcgataaag	gaaaggctat	cggtcaagat	gcctctgccc	acagtgggtc	caaagatgga	6120
ccccaccca	cgaggagcat	cgtggaaaaa	gaagacgttc	caaccacgtc	ttcaaagcaa	6180
gtggattgat	gtgataacat	gggtggagcac	gacactctcg	tctactccaa	gaatatcaaa	6240
gatacagtct	cagaagacca	aagggctatt	gagacttttc	aacaaagggt	aatatcggga	6300
aacctcctcg	gattccattg	cccagctatc	tgtcacttca	tcaaaaggac	agtagaaaag	6360
gaaggtggca	cctacaaaatg	ccatcattgc	gataaaggaa	aggctatcgt	tcaagatgcc	6420
tctgcccaga	gtggtcccaa	agatggaccc	ccaccacga	ggagcatcgt	ggaaaaagaa	6480
gacgttccaa	ccacgtcttc	aaagcaagtg	gattgatgtg	atatctccac	tgacgtaagg	6540
gatgacgcac	aatcccacta	tccttcgcaa	gacctctctc	tatataagga	agttcatttc	6600
atttgagag	gacacgctga	aatcaccagt	ctctctctac	aaatctatct	ctctcgagct	6660
ttcgagatc	atgagatatg	atgagatatg	aaaaagcctg	aactcaccgc	gacgtctgtc	6720
gagaagtttc	tgatcgaaaa	gttcgacagc	gtctccgacc	tgatgcagct	ctcggagggc	6780
gaagaatctc	gtgctttcag	cttcgatgta	ggaggcgctg	gatatgtcct	gcgggtaaat	6840
agctgcgcgc	atgggtttcta	caaagatcgt	tatgtttatc	ggcactttgc	atcggcgccg	6900
ctcccgatcc	cggaagtgct	tgacattggg	gagtttagcg	agagcctgac	ctattgcata	6960
ttcccgccgtg	cacagggtgt	cacgttgcaa	gacctgcctg	aaaccgaact	gcccgcgtgt	7020
ctacaacccg	tcgcggaggc	tatggatgcg	atcgctgcgg	ccgatcttag	ccagacgagc	7080
gggttcggcc	cgggggggca	gcaaggaatc	ggtcaataca	ctacatggcg	tgatttcata	7140
tgccgcgattg	ctgatcccca	tgtgtatcac	tgccaaactg	tgatggacga	caccgtcagt	7200
gcgtccgctc	cgcaggctct	cgatgagctg	atgctttggg	ccgaggactg	ccccgaagtc	7260
cggcacctcg	tgcacgcgga	tttcggctcc	aacaatgtcc	tgacggacaa	tgcccgcata	7320
acagcgggtc	ttgactggag	cgaggcgatg	ttcggggatt	cccaatacga	gggtcgccaac	7380
atcttcttct	ggaggccgtg	gttggcttgt	atggagcagc	agacgcgcta	cttcgagcgg	7440
aggcatccgg	agcttgacgg	atcgccacga	ctccggggcg	atatgtctcg	cattgggtctt	7500
gaccaactct	atcagagctt	ggttgacggc	aatttccgatg	atgcagcttg	ggcgaggggt	7560
cgatgcgacg	caatcgtccg	atccggagcc	gggactgtcg	ggcgtagaca	aatcgcccg	7620
agaagcgccg	ccgtctggac	cgatggctgt	gtagaagtac	tcgccgatag	tggaacccga	7680
cgccccagca	ctcgtccgag	ggcaaaagaa	tagagtagat	gccgaccgga	tctgtcgatc	7740
gacaagctcg	agtttctcca	taataatgtg	tgagtagttc	ccagataagg	gaattaggg	7800
tcctataggg	tttcgctcat	gtgttgagca	tataagaaac	ccttagtatg	tatttgatt	7860
tgtaaaatc	ttctatcaat	aaaattttcta	attcctaaaa	ccaaaatcca	gtactaaaat	7920
ccagatcccc	cgaattaatt	cggcggttaat	tcagatcaag	cttggcactg	gccgtcgttt	7980
tacaacgtcg	tgactgggaa	aaccttggcg	ttacccaact	taatcgctt	gcagcacatc	8040
cccttttcgc	cagctggcgt	aatagcgga	aggcccgac	cgatcgccct	ttcccaacag	8100
tgccgcagcc	gaatggcgaa	tgctagagca	gcttgagctt	ggatcagatt	gtcgtttccc	8160
gccttcagtt	tggggatcct	ctagactgaa	ggcgggaaac	gacaatctga	tcattgagcgg	8220
agaattaagg	gagtcacgtt	atgacccccg	ccgatgacgc	gggacaagcc	gttttacgtt	8280
tggaaactgac	agaaccgcaa	cgttgaaagga	gccactcagc	cgcggttttc	tggaagttta	8340
tgagctaagc	acatacgtca	gaaaccatta	ttgcgcggtc	aaaagtccgc	taaggctcact	8400
atcagctagc	aaatatctct	tgtcaaaaat	gctccactga	cgttccataa	attcccctcg	8460
gtatccaatt	agagtctcat	attcactctc	aatccaaata	atctgcaccg	gatctcgaga	8520
atcgaattcc	cgcggcgcc	atggtagatc	tgactagtaa	aggagaagaa	cttttctgtca	8580
gagttgtccc	aattcttggt	gaattagatg	gtgatgttaa	tgggcacaaa	ttttctgtca	8640
gtggagaggg	tgaagggtgat	gcaacatacg	gaaaacttac	ccttaaatctt	atttgacta	8700
ctggaaaaact	acctgttccg	tggccaacac	ttgtcactac	tttctcttat	ggtgttcaat	8760
gcttttcaag	ataccagat	catatgaagc	ggcagcagct	cttcaagagc	gccatgcctg	8820
agggatcagt	gcaggagagg	accatcttct	tcaaggacga	cggaactac	aagacacgtg	8880
ctgaagtcaa	gtttgaggg	gacaccctcg	tcaacaggat	cgagcttaag	ggaatcgatt	8940

tcaaggaggga	cggaaacatc	ctcggccaca	agttggaata	caactacaac	tcccacaacg	9000
tatacatcat	ggccgacaag	caaaagaacg	gcatcaaagc	caacttcaag	acccgcccaca	9060
acatcgaaga	cggcggcgtg	caactcgctg	atcattatca	acaaaatact	ccaattggcg	9120
atggccctgt	cctttttacca	gacaaccatt	acctgtccac	acaatctgcc	ctttcgaaaag	9180
atcccaacga	aaagagagac	cacatgggtcc	ttcttgagtt	tgtaacagct	gctgggatta	9240
cacatggcat	ggatgaaata	tacaaaagta	gccaccacca	ccaccaccac	gtgtgaattg	9300
gtgaccagct	cgaatttccc	cgatcgttca	aacatttggc	aataaagttt	cttaagattg	9360
aatcctgttg	cgggtcttgc	gatgattatc	atataatttc	tggtgaatta	cgtaaagcat	9420
gtaataatta	acatgtaatg	catgacgtta	tttatgagat	gggtttttat	gattagagtc	9480
ccgcaattat	acatttaata	cgcatagaa	aacaaaatat	agcgcgcaaa	ctaggataaa	9540
ttatcgcgcg	cgggtgtcatc	tatgttacta	gatcggaat	taaactatca	gtgtttgaca	9600
ggatatattg	gcggttaaac	ctaagagaaa	agagcgttta	ttagaataac	ggatatatta	9660
aaggcggtga	aaaggtttat	ccgttcgtcc	atttgatgt	gcatgccaac	cacagggttc	9720
ccctcgggat	caaagtactt	tgatccaacc	cctccgctgc	tatagtgcag	tccggttctg	9780
acgttcagtg	cagccgtctt	ctgaaaacga	catgtcgcac	aagtcctaag	ttacgcgaca	9840
ggctgccgcg	ctgccctttt	cctggcgttt	tcttgcgcg	tggttttagtc	gcataaagta	9900
gaatacttgc	gactagaacc	ggagacatta	cgccatgaac	aagagcgccg	ccgctggcct	9960
gctgggctat	gcccgcgtca	gcaccgacga	ccaggacttg	accaaccaac	gggcccgaact	10020
gcacgcggcc	ggctgcacca	agctgttttc	cgagaagatc	accggcacca	ggcgcgaccg	10080
cccgagctg	ggcaggatgc	ttgaaccact	acgcccggc	gacgttgtga	cagtgcaccg	10140
gctagaccgc	ctggcccgcga	gcacccgcga	cctactggac	attgcccagc	gcatccaggga	10200
ggccgcgcg	ggcctgcgta	gcctggcaga	cccgtgggcc	gacaccacca	cgccggcccg	10260
ccgcatggtg	ttgaccgtgt	tcgcggcgat	tgcccagttc	gagcgttccc	taatcatcga	10320
cgcaccccg	atggggcgcg	aggccgcgca	ggcccgaggg	gtgaagtttg	gcccccgccc	10380
taccctcacc	ccggcacaga	tcgcgcacgc	ccgcgagctg	atcgaccagg	aaggcccgac	10440
cgtgaaagag	gcccgtgcac	tgcttggcgt	gcacgcctcg	accctgtacc	gcgcacttga	10500
gcgcagcgag	aaagtgcgcg	ccaccgaggg	caggcgcgcg	ggtgccttcc	gtgaggacgc	10560
attgaccgag	gcccagcccc	tgccggccgc	cgagaatgaa	cgccaagagg	aacaagcatg	10620
aaaccgcacc	aggacggcca	ggacgaaccg	tttttcatta	ccgaagagat	cgaggcgagg	10680
atgatcgcg	ccgggtacgt	gttcgagcgc	cccgcgcacg	tctcaaccgt	gcccgtgcat	10740
gaaatcctgg	ccggtttgtc	tgatgccaa	ctggcgccct	ggccggccag	cttggccgct	10800
gaagaaaccg	agcgcgcgcg	tctaaaaagg	tgatgtgtat	ttgagtaaaa	cagcttgctg	10860
catgcggtcg	ctgcgtatat	gatgcatga	gtaaataaac	aaatacgcaa	ggggaacgca	10920
tgaaggttat	aaccagaaag	aaccagaaag	gcgggtcagg	caagacgacc	atcgcaaccc	10980
atctagcccc	cgccctgcaa	ctcgccgggg	ccgatgttct	gttagtcgat	tccgatcccc	11040
agggcagtg	ccgcgattgg	gcggccgctg	gggaagatca	accgctaacc	gttgtcgga	11100
tcgaccgccc	gacgattgac	cgcgacgtga	agggccatcg	ccggcgcgac	ttcgtagtga	11160
tcgacggagc	gccccaggcg	gcggacttgg	ctgtgtccgc	gatcaaggca	gccgacttcg	11220
tgctgattcc	ggtgcagcca	agcccttacg	acatatgggc	caccgcccag	ctgggtggagc	11280
tggttaagca	gcccattgag	gtcacggatg	gaaggctaca	agcggccttt	gtcgtgtcgc	11340
ggcgatcaa	aggcacgcgc	atcgccgggt	aggttgccga	ggcgctggcc	gggtacgagc	11400
tgcccatctc	tgagtcccgt	atcacgcagc	gcgtgagcta	cccaggcact	gcccgcgccc	11460
gcacaaccgt	tcttgaatca	gaacccgagg	gcgacgctgc	ccgcgaggtc	caggcgctgg	11520
ccgctgaaat	ttaatacaaaa	ctcatttgag	ttaatgaggt	aaagagaaaa	tgagcaaaaag	11580
cacaaacacg	ctaagtgcgc	gccgtccgag	cgcacgcagc	agcaaggctg	caacgttggc	11640
cagcctggca	gacacgccag	ccatgaagcg	ggtcaacttt	cagttgcccg	cggaggatca	11700
caccaagctg	aagatgtacg	cggtacgcca	aggcaagacc	attaccgagc	tgctatctga	11760
atacatcgcg	cagctaccag	agtaaatgag	caaatagaata	aatgagtaga	tgaatttttag	11820
cggctaaagg	aggcggcgatg	gaaaatcaag	aacaaccagg	caccgacgcc	gtggaatgcc	11880
ccatgtgtgg	aggaacgggc	ggttggccag	gcgtaaagcg	ctgggttgtc	tgccggccct	11940
gcaatggcac	tggaaccccc	aagcccaggg	aatcggcgtg	acggtcgcaa	accatccggc	12000
ccggtacaaa	tcggcgcgcc	gctgggtgat	gacctgggtg	agaagttgaa	ggccgcgcag	12060
gccgcccagc	ggcaacgcat	cgaggcagaa	gcacgccccg	gtgaatcgtg	gcaagcgggc	12120
gctgatcgaa	tccgcaaaga	atcccggcaa	ccgcccggcag	ccggtgcgcc	gtcgattagg	12180
aagccgcccc	agggcgacga	gcaaccagat	tttttcgttc	cgatgctcta	tgacgtgggc	12240
acccgcgata	gtcgcagcat	catggacgtg	gccgttttcc	gtctgtcgaa	gcgtgaccga	12300
cgagctggcg	aggtgatccg	ctacgagctt	ccagacgggc	acgtagaggt	ttccgcaggg	12360
ccggccggca	tgggcagtg	gtgggattac	gacctggtac	tgatggcggt	ttcccatcta	12420
accgaatcca	tgaaccgata	ccgggaaggg	aaggagagaca	agcccggccg	cgtgttccgt	12480
ccacacgttg	cggacgtact	caagttctgc	cggcgagccg	atggcggaag	gcagaaagac	12540
gacctggtag	aaacctgcat	tcggttaaac	accacgcacg	ttgccatgca	gc	12592

<210> 96
 <211> 3357
 <212> DNA
 <213> Artificial Sequence

<220>

<223> pGEMEasyNOS Plasmid

<400> 96

```
tatcactagt gaattcgcgg ccgcctgcag gtgcaccata tgggagagct cccaacgcgt 60
tggatgcata gcttgagtat tctatagtgt cacctaaata gcttggcgta atcatgggtca 120
tagctgtttc ctgtgtgaaa ttgttatccg ctcaacaattc cacacaacat acgagccgga 180
agcataaaagt gtaaaagcctg ggggtgcctaa tgagtgcagct aactcacatt aattgcgttg 240
cgctcactgc ccgctttcca gtccgggaaac ctgtcgtgcc agctgcatta atgaatcggc 300
caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc gctcactgac 360
tcgctgcgct cggctcgttcg gctgcggcga gcggtatcag ctcaactcaa ggcggtaata 420
cggttatcca cagaatcagg ggataacgca ggaagaaca tgtgagcaa aggccagcaa 480
aagccaggga accgtaaaaa ggccgcgttg ctggcggttt tccataggct ccgccccct 540
gacgagcatc acaaaaatcg acgctcaagt cagaggtggc gaaaccgac aggactataa 600
agataccagg cgtttccccc tggaaagctcc ctctgttcc gacctgccc 660
cttaccggat acctgtccgc ggtgtaggtc gtgcgctta tccggttaact atcgctctga gtccaaccg 840
cgctgtagggt atctcagttc ctgcgcctta gccactggta acaggattag cagagcgagg 900
cccccgcttc agcccgaccg acttatcgcc actggcagca tccggttaact actacggcta cactagaaga 960
gtaagacacg acttatcgcc actggcagca tccggttaact actacggcta cactagaaga 1020
tatgtaggcg gtgctacaga gttcctgaag gatccttga agctgggctg tgtgcacgaa 780
acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc 1020
tcttgatccg gcaaacaaac caccgctggg agcggtggtt tttttgtttg caagcagcag 1080
attacgcgca gaaaaaaagg acgttaaggg atttttgtca tgagattatc caatctaaag tatatatgag 1260
gctcagtgga acgaaaaactc ttaaaaaatga agttttaaat caatctaaag tatatatgag 1260
ttcacctaga tctttttaa ccaatgctta atcagtgagg cacctatctc agcgatctgt 1320
taaacttggt ctgacagtta tgccctgactc cccgtcgtgt agataactac gatacgggag 1380
ctatttcggt catccatagt tgcctgactc ataccgcgag acccacgctc accggctcca 1440
ggcttaccat ctggccccag gccagccgga agggccgagc gcagaagtgg tcctgcaact 1500
gatttatcag caataaacca tattaattgt tggtgccatt ctccggttcc tagtctctc 1560
ttatccgctt ccacccagtc tggtgccatt ctccggttcc tagtctctc 1560
gttaatatgt ttctattcag ctccggttcc tagtctctc 1560
atgttgtgca aaaaagcggg tagctccttc ggtcctccga tcgttgtcag aagtaagttg 1740
gccgcagtgt tatcactcat ggttatggca ggttatggca attctcttac tgtcatgcca 1800
tccgtaagat gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt 1860
atgcccgcag cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc 1920
agaactttaa agagtctcat cattggaaaa cgttcttcgg ggcgaaaaact ctcaaggatc 1980
ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg caccacaactg atcttcagca 2040
tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa 2100
aagggataaa gggcgacacg gaaatgttga atactatac tcttctttt tcaatattat 2160
tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa 2220
aataaacaaa taggggttcc gcgcacattt ccccgaaaag tgccacctga tgccggtgta 2280
aatccgcac agatgcgtaa ggagaaaaa cccgcacagg aaattgtaag cgttaatatt 2340
ttgttaaaat tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa 2400
atcggaacaaa tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca 2460
gtttggaaca agagtccact attaaagaac gtggactcca acgtcaagg gcgaaaaaac 2520
gtctatcagg gcgatggccc actacgtgaa ccatcacctt aatcaagttt tttggggtcg 2580
aggtgccgta aagcactaaa tcggaaccct aaaggagacc cccgatttag agcttgacgg 2640
ggaaagccgg cgaacgtggc gagaaaggaa gggaaggaa cgaaggagc gggcgctagg 2700
gcgctggcaa gtgtagcggg cacgctgcgc gtaaccacca caccgcgcg gcttaatgag 2760
ccgctacagg gcgcgtccat tcgccattca ggctgcgcaa ctggtgggaa gggcgatcgg 2820
tgccggcctc ttcgctatta cgccagctgg cgaaagggg atgtgctgca aggcgattaa 2880
gttggttaac gccagggttt tcccagtcac gacgttgtaa aacgacggc agtgaattgt 2940
aatacgaact actatagggc gaattgggccc cgactcggc cgccatggcg 3000
gccgcgggaa ttcgattctc gagatccggg gcagattatt tggattgaga gtgaatatga 3060
gactctaatt ggataccgag ggaatttat ggaacgtcag tggagcattt ttgacaagaa 3120
atatttgcta gctgatagtg accttaggcg ccttttgaa ggcgaataat ggtttctgac 3180
gtatgtgctt agctcattaa actccagaaa cccgcggctg agtggctcct tcaacgtttg 3240
ggttctgtca gttccaaacg taaaacggct tgtccgcgt catcgccggg ggtcataacg 3300
tgactccctt aattctccgc tcatgatcag attgtcgttt cccgccttca gtctaga 3357
```

<210> 97

<211> 10122

<212> DNA

<213> Artificial Sequence

<220>

<223> p1302NOS Plasmid

<400> 97

catggtagat	ctgactagta	aaggagaaga	acttttctact	ggagttgtcc	caattcttctgt	60
tgaattagat	gggtgatgta	atgggcacaa	attttctgtc	agtggagagg	gtgaagggtga	120
tgcaacatac	ggaaaactta	cccttaaat	tatttgcact	actggaaaac	tacctgttcc	180
gtggccaaca	cttgtcacta	ctttctctta	tgggtgtcaa	tgcttttcaa	gataccacaga	240
tcatatgaag	cggcaccgact	tcttcaagag	cgccatgcct	gagggatagc	tgcaggagag	300
gaccatcttc	ttcaaggacg	acgggaacta	caagacacgt	gctgaagtca	agtgtgaggg	360
agacaccctc	gtcaacagga	tcgagcttaa	gggaatcgat	ttcaaggagg	acggaaacat	420
cctcgccac	aagttggaat	acaactacaa	ctcccacaa	gtatacatca	tggccgacaa	480
gcaaaagaac	ggcatcaaag	ccaacttcaa	gacccgccac	aacatcgaag	acggcggcgt	540
gcaactcgct	gacattatc	aacaaaatac	tccaattggc	gatggccctg	tccttttacc	600
agacaaccat	tacctgtcca	cacaatctgc	cctttcgaaa	gatcccaacg	aaaagagaga	660
ccacatgggtc	cttcttgagt	ttgtaacagc	tgctgggatt	acacatggca	tggatgaact	720
atacaaagct	aaacatttgg	accaccacca	acgtgtaatt	ggtgaccagc	tcgaatttcc	780
ccgatcggtc	catataat	caataaagtt	tccttaagatt	gaatcctgtt	gccggtcttg	840
cgatgattat	catataat	ctggttgaatt	acgttaagca	tgtataaatt	aacatgtaat	900
gcatgacgtt	atttatgaga	tgggttttta	tgattagagt	cccgaatta	tacattta	960
acgcgataga	aaacaaaata	tagcgcgcaa	actaggataa	attatcgccg	gcggtgtcat	1020
ctatgttact	agatcgggaa	ttaaactatc	agtgtttgac	aggatatatt	ggcgggtaaa	1080
cctaagagaa	aagagcgttt	attagaataa	cggtatatta	aaagggcgtg	aaaaggttta	1140
tccgttcgtc	catttgtatg	tgcattgcaa	ccacagggtt	cccctcggga	tcaaagtact	1200
ttgatccaac	ccctccgctg	ctatagtga	gtcggcttct	gacgttcagt	gcagccgtct	1260
tctgaaaacg	acatgtcgca	caagtcctaa	gttacgcgac	aggctgccc	cctgcccttt	1320
tcctggcggtt	ttcttgtcgc	gtgttttagt	cgcataaagt	agaatacttg	cgactagaac	1380
cggagacatt	agcccatgaa	caagagcgcc	gcccgtggcc	tgctgggcta	tgcccgctc	1440
agcaccgacg	accaggactt	gaccaaccaa	cgggcccgaac	tgcacgcggc	cggctgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccttgg	cgacgttgtg	acagtacca	ggctagaccg	cctggcccgc	1620
agcaccgcg	acctactgga	cattgcccag	cgcatccagg	aggcggcgcc	gggctcgct	1680
agcctggcag	agccgtgggc	cgacaccacc	acgcccggcg	gcccgcattg	gttgaccgtg	1740
ttcgccggca	ttgcccaggt	cgagcgttcc	ctaactcatc	accgcaccgc	gagcgggccc	1800
gagcccgcca	agccccgagg	cgtagaagtt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgccgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggctgca	1920
ctgcttggcg	tgcattcgctc	gaccctgtac	cgcgcaactg	agcgcagcga	ggaagtga	1980
cccaccgagg	cagggcggcg	cggtgccttc	cgtaggagcg	cattgaccga	ggccgacgcc	2040
ctggccggcg	ccggaagatg	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gttttttcatt	accgaagaga	tcgaggcgga	gatgatcgcg	gcccgggtacg	2160
tgcttcgagcc	gcccgcgcac	gtctcaaccg	tgccggctgca	tgaatcctg	gcccgtttgt	2220
ctgatgccaa	gctggcgccc	tgccgggcca	gcttggccgc	tgaagaaacc	gagcggccgc	2280
gtctaaaaag	gtgatgtgta	tttgagtaaa	acagcttgcg	tcatgcggctc	gctgcgtata	2340
tgatcgcatg	agtaataaaa	caaatacgca	aggggaaacg	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcgcccgcga	2460
actcgccggg	gcccgatgttc	tgtagtgcga	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcggccggtg	cgggaagatc	aaccgcctaac	cgttgtcggc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gccggcgcca	cttcgtagtg	atcgacggag	cgccccaggc	2640
ggcggacttg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cgggtgcagcc	2700
aagcccttac	gacatatggg	ccaccgcoga	cctggtggag	ctggttaagc	agcgcattga	2760
ggtcacggat	ggaaggctac	aagcggcctt	tgtcgtgtcg	cgggcatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	aggcgtgtgc	cgggtacgag	ctgcccattc	ttgagtcgcc	2880
tatcacgcag	cgctgtgagc	acccaggcac	tgccgcggcc	ggcacaaccg	ttcttgaatc	2940
agaaccgcag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gcccgtgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgcc	3060
ggcgtccga	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcga	3120
gccatgaagc	gggtcaactt	tcagttgccc	gcggaggatc	acaccaagct	gaagatgtac	3180
gcggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcggctaaa	gagggcgcat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaaacggg	3360
cggttggcca	ggcgtaagcg	gctgggttgt	ctgccggccc	tgcaatggca	ctggaaacccc	3420
caagcccag	gaatcggcgt	gacgggtcga	aaccatccgg	cccgtacaa	atcgccgcgg	3480
cgctgggtga	tgacctgggtg	gagaagttga	aggccgcgca	ggccgcccag	cggaacacga	3540
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcggc	cgctgatcga	atccgcaaa	3600
aatcccggca	accgcccggca	gccgggtgccc	cgctcgattag	gaagccgccc	aagggcgacg	3660
agcaaccaga	ttttttcgtt	ccgatgctct	atgacgtggg	caccgcgat	agtcgcagca	3720
tcatggacgt	ggccggttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctaccgagct	tccagacggg	cacgtagagg	tttccgcagg	gcccggccggc	atggccagtg	3840

tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaaggagagac	aagcccggcc	gcgtgtttccg	tccacacgtt	gcggacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaga	cgacctggta	gaaacctgca	4020
ttcggttaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccggcc	4080
tggtagcgg	atccgaggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcccgg	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccgacgtg	ctgacggttc	accccgatta	ctttttgatc	gatcccggca	4260
tcggccggtt	tctctaccgc	ctggcacgcc	gcgccgcagg	caaggcagaa	gccagatggt	4320
tgttcaagac	gatctacgaa	cgcatgtggca	gcgccggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
cgggttccta	atgtacggag	cagatgctag	ggcaaaattgc	cctagcaggg	gaaaaaggtc	4560
gaaaagggtc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtaacattgg	gaaccgggtca	cacatgtaag	4680
tgactgatata	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aaccgcctg	gcctgtgcat	aactgtctgg	ccagcgcaaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggctcgtgc	gctccctacg	ccccgccgct	tcgcgtcggc	4860
ctatcgccgg	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcggacaagc	cgccgcgctc	ccactcgacc	gccggcgccc	acatcaaggc	accctgcctc	4980
gcgctgttcc	gtgatcgagg	tgaaaacctc	tgacacatgc	agctcccggg	gacgggtcaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgct	agcgggtggt	5100
ggcgggtgtc	ggggcgagc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgc	ggcatcagag	cagattgtac	tgagatgca	ccatatgcgg	tgtgaaatac	5220
cgcacagatg	gctaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgtcgcg	ctcggctcgt	cggctgcggc	gagcggatc	agctcactca	aaggcggtaa	5340
tacggttatc	cacagaatca	ggggataacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	gcgaaaaccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgtcgc	ctctcctggt	ccgaccctgc	5580
cgcttacggg	ataccgttcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aaccccccg	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgtcct	gagtccaacc	5760
cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggtagctagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggcaaaaca	accaccgctg	gtagcgggtg	tttttttggt	tgcaagcagc	6000
agattacggc	cagaaaaaaa	ggatctcaag	cttccggcctt	gatcttttct	acgggggtctg	6060
acgctcagtg	gaacgaaaac	tcacgttaag	ggattttggt	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaata	taatatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatata	ctccctgatc	gaccggagcg	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaaggccac	6300
ttactttgcc	atctttcaca	aagatgttgc	tgtctcccag	gtcgccgctg	gaaaagacaa	6360
gttctcttcc	gggcttttcc	gtcttttaaa	aatcatacag	ctcgccggga	tctttaaatg	6420
gagtgctctc	ttcccagttt	tcgcaatcca	catcgccag	atcgttattc	agtaagtaac	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tccgcataca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgctccagc	6660
catcatgccc	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcct	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttcattt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttcga	tcagtttttt	caattccggt	gatattctca	6900
ttttagccat	ttattatttc	cttctctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaaat	ccaattcact	gttccttgca	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	ttggcgctat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
tacctccgc	gagatcatcc	gtgtttcaaa	cccgccagct	tagttgccgt	tcttccgaat	7200
agcatcggt	acatgagcaa	agtctgccgc	cttacaacgg	ctctcccgct	gacgcgctcc	7260
cggactgatg	ggctgcctgt	atcgagtgg	gatttttgtg	cgagctgccg	gtcggggagc	7320
tgttggtcgg	ctggtggcag	gatataattg	ggtgtaaaac	aattgacgct	tagacaactt	7380
aataacacat	tgccgacggt	tttaattgtac	tgaattaaac	ccgaattaat	tcgggggatc	7440
tggattttag	tactggattt	tggttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgctcaa	cacatgagcg	aaaccctata	7560
ggaaccctaa	ttcccttatc	tgggaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccgggt	ggcatctact	ctatttcttt	gccctcggac	gagtgtcggg	7680
gcgtcgggtt	ccactatcgg	cgagtacttc	tcacagccca	tcggctcaga	cggccgcgct	7740
tctgcggggc	atgtgtgtac	gcccgcagct	cccggtccg	gatcggacga	ttgcgtcgca	7800
tcgacctgc	gcccagctg	catcatcgaa	attgcgctca	accaagctct	gatagagttg	7860

```

gtcaagacca atgctggagca tatacgcccc gagtcgtggc gatcctgcaa gctccggatg 7920
ctcccgctcg aagtagcgcg tctgctgctc catacaagcc aaccacggcc tccagaagaa 7980
gatgttggcg acctcgattt gggaatcccc gaacatcgcc tcgctccagt caatgacggc 8040
tgttatgcgg ccattgtccg tcaggacatt gttggagccg aaatccgcgt gcacgaggtg 8100
ccggacttcg gggcagtcct cggcccaaag catcagctca tcgagagcct gcgcgacgga 8160
cgactgacg gtgtcgctcca tcacagtttg ccagtgtata acatggggat cagcaatcgc 8220
gcatatgaaa tcacgccatg tagtgtattg accgattcct tgcgggtccga atgggcccga 8280
cccgcctcgc tggctaagat cggccgcagc gatcgcatcc atagcctccg cgaccgggtg 8340
tagaacagcg ggcagttcgg ttccaggcag gtcttgcaac gtgacaccct gtgcacggcg 8400
ggagatgcaa taggtcaggg tctcgctaaa ctccccaatg tcaagcactt ccggaatcgg 8460
gagcgcgggc gatgcaaagt gccgataaac ataacgatct ttgtagaaac catcggcgca 8520
gctatttacc cgcaggacat atccacgccc tcctacatcg aagctgaaag cagagattc 8580
ttcgccctcc gagagctgca tcagggtcggg gacgtgtcgt aacttttctga tcagaaactt 8640
ctcgacagac gtcgcgggtg gtccaggctt tttcatatct cattgcccc ccggatctgc 8700
gaaagctcga gagagataga tttgtagaga gagactgggt atttcagcgt gtcctctcca 8760
aatgaaaatga acttccttat atagaggaag gtcttgcgaa ggatagtggg attgtgcgtc 8820
atcccttacg tcagtggaga tatcacatca atccacttgc tttgaagacg tgggtggaac 8880
gtcttctttt tccacgatgc tctcgtggg tgggggtcca tctttgggac cactgtcggc 8940
agaggcatct tgaacgatag cctttccttt atcgcaatga tggcatttgt aggtgccacc 9000
ttccttttct actgtccttt tgatgaagtg acagatagct gggcaatgga atccgaggag 9060
gtttcccgat attacccttt gttgaaaagt ctcaatagcc ctttgggtct ctgagactgt 9120
atctttgata ttcttgaggt agacgagagt gtcgtgctcc accatgttat cacatcaatc 9180
cacttgcttt gaagacgtgg ttggaacgtc tcttttttcc acgatgctcc tcgtgggtgg 9240
gggtccatct ttgggaccac tgtcggcaga ggcattctga acgatagcct ttcttttatc 9300
gcaatgatgg catttgtagg tgccaccttc cttttctact gtccttttga tgaagtgaca 9360
gatagctggg caatggaatc cgaggaggtt tcccgatatt accctttgtt gaaaagtctc 9420
aatgacccct tggctctctg agactgtatc tttgatattc ttggagtga cgagagtgtc 9480
gtgctccacc atgttggtgaa gctgctctag ccaatacgca aaccgcctct ccccgcgctg 9540
tggccgattc ataatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag 9600
cgcaacgcaa ttaatgtgag ttagctcact cattaggcac ccaggctttt acactttatg 9660
cttccggctc gtatgttgtg tgggaattgt agcggataac aatttcacac aggaaacagc 9720
tatgaccatg attatgaatt cgagctcggg acccggggat cctctagact gaaggcggga 9780
aacgcgaatc ttagctatgag cggagaatta agggagtac gttatgacc ccgccgtga 9840
cgcggggaaa gccgttttac gtttggaaact gacagaaccg caacgttgaa ggagccactc 9900
agcccggggt ttctggagtt taatgagcta agcacatacg tcagaaacca ttattgcgcg 9960
ttcaaaagtc gcctaagggt actatcagct agcaaatatt tcttgtcaaa aatgctccac 10020
tgacgtttcca taaattcccc tcggtatcca attagagtct catattcact ctcaatccaa 10080
ataatctgca ccggatctcg agaatcgaat tcccgcggcc gc 10122

```

<210> 98

<211> 621

<212> DNA

<213> Artificial Sequence

<220>

<223> N. tabacum rDNA intergenic spacer (IGS) sequence

<300>

<308> Genbank #Y08422

<309> 1997-10-31

<400> 98

```

gtgctagcca atgtttaaca agatgtcaag cacaatgaat gttggtggtt ggtggtcgtg 60
gctggcggtg gtggaaaatt gcggtggttc gagcggtagt gatcggcgat ggttggtgtt 120
tgcagcggtg tttgatatcg gaatcactta tgggtggttg cacaatggag gtgcgtcatg 180
gttattgggtg gttggtcatc tatatatatt tataataata ttaagtattt tacctatttt 240
ttacatattt ttattataat ttatgcattg tttgtatttt taaatagttt ttatcgtact 300
tgttttataa aatattttat tattttatgt gttatattat tacttgatgt attggaaatt 360
ttctccattg tttttctat atttataata attttcttat tttttttgtt ttattatgt 420
attttttcgt tttataataa atatttatta aaaaaaatat tatttttgta aaatatatca 480
tttacaatgt ttaaaagtca tttgtgaata tattagctaa gttgtacttc tttttgtgca 540
tttggtgttg tacatgtcta ttatgattct ctggccaaaa catgtctact cctgtcactt 600
gggttttttt ttttaagaca t 621

```

<210> 99

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-F1 Primer

<400> 99

gtgctagcca atgtttaaca agatg 25

<210> 100

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-R1 Primer

<400> 100

atgtcttaaa aaaaaaaacc caagtgc 28

<210> 101

<211> 233

<212> DNA

<213> Mus Musculus

<300>

<308> Genbank #V00846

<309> 1989-07-06

<400> 101

gacctggaat atggcgagaa aactgaaaat cacggaaaat gagaaataca cactttagga 60
cgtgaaatat ggcgaggaaa actgaaaaag gtggaaaatt tagaaatgtc cactgtagga 120
cgtggaatat ggcaagaaaa ctgaaaatca tggaaaatga gaaacatcca cttgacgact 180
tgaaaaatga cgaaatcact aaaaaacgtg aaaaatgaga aatgcacact gaa 233

<210> 102

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-F1 Primer

<400> 102

aataccgcgg aagcttgacc tggaatatcg c 31

<210> 103

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-R1 Primer

<400> 103

ataaccgcgg agtccttcag tgtgcat 27

<210> 104

<211> 277

<212> DNA

<213> Artificial Sequence

<220>

<223> Nopaline Synthase Promoter Sequence

<300>

<308> Genbank #U09365

<309> 1997-10-17

<400> 104
gagctcgaat ttccccgcatc gttcaaaccat ttggcaataa agtttcttaa gattgaatcc 60
tggtgcgggt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cggtatttat gagatgggtt tttatgatta gagtcccgca 180
attatacatt taatacgcca tagaaaacaa aatatagcgc gcaaaactagg ataaattatc 240
gcgcgcgggtg tcactctatgt tactagatcg ggaattc 277

<210> 105
<211> 1812
<212> DNA
<213> Escherichia coli

<220>
<221> CDS
<222> (1)...(1812)
<223> Beta-Glucuronidase

<300>
<308> Genbank #S69414
<309> 1994-09-23

<400> 105
atg tta cgt cct gta gaa acc cca acc cgt gaa atc aaa aaa ctc gac 48
Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
1 5 10 15
ggc ctg tgg gca ttc agt ctg gat cgc gaa aac tgt gga att gat cag 96
Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
20 25 30
cgt tgg tgg gaa agc gcg tta caa gaa agc cgg gca att gct gtg cca 144
Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
35 40 45
ggc agt ttt aac gat cag ttc gcc gat gca gat att cgt aat tat gcg 192
Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
50 55 60
ggc aac gtc tgg tat cag cgc gaa gtc ttt ata ccg aaa ggt tgg gca 240
Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
65 70 75 80
ggc cag cgt atc gtg ctg cgt ttc gat gcg gtc act cat tac ggc aaa 288
Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
85 90 95
gtg tgg gtc aat aat cag gaa gtg atg gag cat cag ggc ggc tat acg 336
Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
100 105 110
cca ttt gaa gcc gat gtc acg ccg tat gtt att gcc ggg aaa agt gta 384
Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
115 120 125
cgt atc acc gtt tgt gtg aac aac gaa ctg aac tgg cag act atc ccg 432
Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
130 135 140
ccg gga atg gtg att acc gac gaa aac ggc aag aaa aag cag tct tac 480
Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
145 150 155 160
ttc cat gat ttc ttt aac tat gcc gga atc cat cgc agc gta atg ctc 528
Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
165 170 175
tac acc acg ccg aac acc tgg gtg gac gat atc acc gtg gtg acg cat 576

Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185					190			
gtc	gcg	caa	gac	tgt	aac	cac	gcg	tct	gtt	gac	tgg	cag	gtg	gtg	gcc	624
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala	
		195					200					205				
aat	ggt	gat	gtc	agc	gtt	gaa	ctg	cgt	gat	gcg	gat	caa	cag	gtg	gtt	672
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val	
	210					215					220					
gca	act	gga	caa	ggc	act	agc	ggg	act	ttg	caa	gtg	gtg	aat	ccg	cac	720
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His	
	225				230					235					240	
ctc	tgg	caa	ccg	ggt	gaa	ggt	tat	ctc	tat	gaa	ctg	tgc	gtc	aca	gcc	768
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala	
				245					250					255		
aaa	agc	cag	aca	gag	tgt	gat	atc	tac	ccg	ctt	cgc	gtc	ggc	atc	cgg	816
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg	
			260					265					270			
tca	gtg	gca	gtg	aag	ggc	gaa	cag	ttc	ctg	att	aac	cac	aaa	ccg	ttc	864
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe	
		275					280					285				
tac	ttt	act	ggc	ttt	ggt	cgt	cat	gaa	gat	gcg	gac	ttg	cgt	ggc	aaa	912
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys	
	290					295					300					
gga	ttc	gat	aac	gtg	ctg	atg	gtg	cac	gac	cac	gca	tta	atg	gac	tgg	960
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp	
	305			310					315					320		
att	ggg	gcc	aac	tcc	tac	cgt	acc	tcg	cat	tac	cct	tac	gct	gaa	gag	1008
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu	
			325						330				335			
atg	ctc	gac	tgg	gca	gat	gaa	cat	ggc	atc	gtg	gtg	att	gat	gaa	act	1056
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr	
			340					345					350			
gct	gct	gtc	ggc	ttt	aac	ctc	tct	tta	ggc	att	ggt	ttc	gaa	gcg	ggc	1104
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly	
		355					360					365				
aac	aag	ccg	aaa	gaa	ctg	tac	agc	gaa	gag	gca	gtc	aac	ggg	gaa	act	1152
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr	
	370					375					380					
cag	caa	gcg	cac	tta	cag	gcg	att	aaa	gag	ctg	ata	gcg	cgt	gac	aaa	1200
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys	
				390						395				400		
aac	cac	cca	agc	gtg	gtg	atg	tgg	agt	att	gcc	aac	gaa	ccg	gat	acc	1248
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr	
				405					410					415		
cgt	ccg	caa	ggt	gca	cgg	gaa	tat	ttc	gcg	cca	ctg	gcg	gaa	gca	acg	1296
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr	
			420					425					430			
cgt	aaa	ctc	gac	ccg	acg	cgt	ccg	atc	acc	tgc	gtc	aat	gta	atg	ttc	1344
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe	
		435					440					445				

tgc	gac	gct	cac	acc	gat	acc	atc	agc	gat	ctc	ttt	gat	gtg	ctg	tgc	1392
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys	
	450					455					460					
ctg	aac	cgt	tat	tac	gga	tgg	tat	gtc	caa	agc	ggc	gat	ttg	gaa	acg	1440
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr	
	465				470					475					480	
gca	gag	aag	gta	ctg	gaa	aaa	gaa	ctt	ctg	gcc	tgg	cag	gag	aaa	ctg	1488
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu	
				485					490					495		
cat	cag	ccg	att	atc	atc	acc	gaa	tac	ggc	gtg	gat	acg	tta	gcc	ggg	1536
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly	
			500					505					510			
ctg	cac	tca	atg	tac	acc	gac	atg	tgg	agt	gaa	gag	tat	cag	tgt	gca	1584
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala	
		515				520						525				
tgg	ctg	gat	atg	tat	cac	cgc	gtc	ttt	gat	cgc	gtc	agc	gcc	gtc	gtc	1632
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val	
	530					535					540					
ggt	gaa	cag	gta	tgg	aat	ttc	gcc	gat	ttt	gcg	acc	tcg	caa	ggc	ata	1680
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile	
	545				550					555					560	
ttg	cgc	gtt	ggc	ggt	aac	aag	aaa	ggg	atc	ttc	act	cgc	gac	cgc	aaa	1728
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys	
			565						570					575		
ccg	aag	tcg	gcg	gct	ttt	ctg	ctg	caa	aaa	cgc	tgg	act	ggc	atg	aac	1776
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Lys	Lys	Arg	Trp	Thr	Gly	Met	Asn	
			580					585					590			
ttc	ggt	gaa	aaa	ccg	cag	cag	gga	ggc	aaa	caa	tga					1812
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln	*					
	595						600									

<210> 106
 <211> 603
 <212> PRT
 <213> Escherichia coli

<300>
 <308> Genbank #S69414
 <309> 1994-09-23

<400> 106
 Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
 1 5 10 15
 Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
 20 25 30
 Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
 35 40 45
 Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
 50 55 60
 Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
 65 70 75 80
 Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
 85 90 95
 Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
 100 105 110
 Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
 115 120 125

Arg	Ile	Thr	Val	Cys	Val	Asn	Asn	Glu	Leu	Asn	Trp	Gln	Thr	Ile	Pro
130						135					140				
Pro	Gly	Met	Val	Ile	Thr	Asp	Glu	Asn	Gly	Lys	Lys	Lys	Gln	Ser	Tyr
145						150				155					160
Phe	His	Asp	Phe	Phe	Asn	Tyr	Ala	Gly	Ile	His	Arg	Ser	Val	Met	Leu
					165				170					175	
Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His
			180					185					190		
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala
		195					200					205			
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val
210						215				220					
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His
225					230					235					240
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala
				245					250					255	
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg
			260					265					270		
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe
		275					280					285			
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys
	290					295				300					
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp
305					310					315					320
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu
				325					330					335	
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr
			340					345					350		
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly
		355					360					365			
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr
	370					375				380					
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys
385					390					395					400
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr
				405					410					415	
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr
			420					425					430		
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe
		435					440					445			
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys
	450					455				460					
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr
465					470					475					480
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu
				485					490					495	
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly
		500						505					510		
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala
		515					520					525			
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val
	530					535					540				
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile
545					550					555					560
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys
				565					570					575	
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn
			580					585					590		
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln					
		595					600								

<210> 107

<211> 277

<212> DNA

<213> Artificial Sequence

<220>
<223> Nopaline Synthase Terminator Sequence

<300>
<308> U09365
<309> 1995-10-17

<400> 107
gagctcgaat ttccccgcatc gttcaaacat ttggcaataa agtttcttaa gattgaatcc 60
tggtgcccgt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta ggtccccgca 180
attatacatt taatagcgga tagaaaacaa aatatagcgc gcaaaactagg ataaattatc 240
gcgcgcgggt tcactctatgt tactagatcg ggaattc 277

<210> 108
<211> 3451
<212> DNA
<213> Artificial Sequence

<220>
<223> HindIII Fragment containing the beta-glucuronidase coding sequence, the rDNA intergenic spacer, and the Mast1 sequence

<400> 108
aagcttgacc tgggaatatcg cgagtaaact gaaaatcacg gaaaatgaga aatacacact 60
ttaggacgtg aaatatggcg aggaaaactg aaaaagggtg aaaatttaga aatgtccact 120
gtaggacgtg gaatatggca agaaaactga aaatcatgga aaatgagaaa catccacttg 180
acgacttgaa aaatgacgaa atcactaaaa aacgtgaaaa atgagaaaat cacactgaag 240
gactccgcgg gaattcgatt gtgctagcca atgtttaaca agatgtcaag cacaatgaat 300
gttggtggtt ggtggtcggt gctggcgggt gtggaaaatt gcgggtggtc gagcggtagt 360
gatcggcgat ggttggtggt tgcagcgggt tttgatatcg gaatcactta tgggtggtgt 420
cacaatggag gtgcgctcatg gttattggtg gttggtcatc tatatatatt tataaataata 480
ttaagtattt tacctatttt ttacatattt tttattaaat ttatgcattg tttgtatttt 540
taaatagttt ttatcgactt tgttttataa aatattttat tattttatgt gttatattat 600
tacttgatgt attggaaatt ttctccattg ttttttctat atttataata attttcttat 660
ttttttttgt tttattatgt atttttctgt tttataataa atatttatta aaaaaaatat 720
tatttttgta aaatatatca tttacaatgt ttaaaagtca tttgtgaata tattagctaa 780
gttgactctc tttttgtgca tttggtgttg tacatgtcta ttatgattct ctggccaaaa 840
catgtctact cctgtcactt gggttttttt ttttaagaca taatcactag tgattatatc 900
tagactgaag gcgggaaacg acaatctgat catgagcggg gaattaaggg agtcacgtta 960
tgacccccgc cgatgacgcg ggacaagccg ttttacgttt ggaactgaca gaaccgcaac 1020
gttgaaaggag ccactcagcc gcgggtttct ggagtttaat gagctaagca catacgtcag 1080
aaaccattat tgcgcgttca aaagtcgcct aaggtcacta tcagctagca aatatttctt 1140
gtcaaaaaatg cctccatgac gttccataaa tttccctcgg tatccaatta gagtctcata 1200
ttcactctca atccaaataa tctgcaccgg atctcgagat cgaattcccg cggccgcgaa 1260
ttcactagtg gatccccggg tacggctcagt cccttatggt acgtcctgta gaaaccccaa 1320
cccgtgaaat caaaaaactc gacggcctgt gggcattcag tctggatcgc gaaaactgtg 1380
gaattgagca gcgttggttg gaaagcgcgt tacaagaaag ccgggcaatt gctgtgccag 1440
gcagttttta cgatcagttc gccgatgcag atattcgtaa ttatgtgggc aacgtctggt 1500
atcagcgcga agtctttata aaagtgtggg tcaataatca gcgtatcgtg ctgctgttct 1560
atgcggtcac tcattacggc aaagtgtggg ggcaggcca ggaagtgtg gagcatcagg 1620
gcggtctata gccatttgaa gccgatgtca cgccgtatgt tattgccggg aaaagtgtac 1680
gtatcacagt ttgtgtgaac aacgaactga actggcagac tatcccgcgg ggaatggtga 1740
ttaccgacga aaacggcaag aaaaagcagt ctacttcca tgatttcttt aactacgccc 1800
ggatccatcg cagcgtaatg ctctacacca cgccgaacac ctgggtggac gatatacccg 1860
tggtgacgca tgtcgcgcaa gactgtaacc acgcgtctgt tgactggcag gtggtggcca 1920
atggtgatgt cagcgttgaa ctgctgatg cggatcaaca actggacaag 1980
gcaccagcgg gactttgcaa gtggtgaatc cgcacctctg gcaaccgggt gaaggttatc 2040
tctatgaact gtacgtcaca gccaaaagcc agacagagtg tgatatctac ccgctgcgcg 2100
tcggcatccg gtcagtggca gtgaagggcg aacagttcct gatcaaccac aaaccgttct 2160
actttactgg ctttgccgtg catgaagatg cggatttgcg cggcaaaagg ttcgataacg 2220
tgctgatggt gcacgatcac gcattaatgg actggatgg ggccaactcc taccgtacct 2280
cgcattaccc ttacgctgaa gagatgctgc actgggcaga tgaacatggc atcgtggtga 2340
ttgatgaaac tgcagctgtc ggctttaacc tctcttagg cattggtttc gaagcggcca 2400
acaagccgaa tgaactgtac agcgaagagg cagtcaacgg ggaaactcag caggcgcact 2460
tacaggcgat taaagagctg atagcgcgtg acaaaaacca cccaagcgtg gtgatgtgga 2520

```
gtattgccaa cgaaccggat acccgtccgc aaggtgcacg ggaatatttc gcgccactgg 2580
cggaagcaac gcgtaaaactc gatccgacgc gtccgatcac ctgcgtcaat gtaatgttct 2640
gcgacgctca caccgatacc atcagcgatc tctttgatgt gctgtgcctg aaccggtatt 2700
acggttggta tgtccaaagc ggcgatttgg aaacggcaga gaaggtactg gaaaaagaac 2760
ttctggcctg gcaggagaaa ctgcatcagc cgattatcat caccgaatac ggcgtggata 2820
cgtagccgg gctgcactca atgtacaccg acatgtggag tgaagagtat cagtgtgcat 2880
ggctggatat gstatcaccgc gtctttgatc gcgtcagcgc cgtcgtcggc gaacagggtat 2940
ggaatttcgc cgatttttgc acctcgcaag gcatattgcg cgttggcggc aacaagaagg 3000
ggatcttcac ccgcgaccgc aaacogaagt cggcggtttt tctgctgcaa aaacgctgga 3060
ctggcatgaa cttcgggtgaa agggaggcaa acaatgaatc aacaactctc 3120
ctggcgacc atcgtcggct acagcctcgg gaattgcgta ccgagctcga atttccccga 3180
tcgttcaaac atttggcaat aaagtttctt aagattgaat cctgttgccg gtcttgccat 3240
gattatcata taatttctgt tgaattacgt taagcatgta ataattaaca tgtaatgcat 3300
gacgttattt atgagatggg tttttatgat tagagtcgcc caattataca tttaatacgc 3360
gatagaaaac aaaatatagc gcgcaaaacta ggataaatta tcgcgcgcgg tgatcatctat 3420
gttactagat cgggaattcg atatcaagct t 3451
```

<210> 109

<211> 14627

<212> DNA

<213> Artificial Sequence

<220>

<223> pAg11a Plasmid

<400> 109

```
catgccaacc acagggttcc cctcgggata aaagtacttt gatccaaccc ctccgctgct 60
atagtgcagt cggcttctga cgttcagtg cagccgtcttc tgaaaacgac atgtcgcaca 120
agtcctaagt tacgcgacag gctgcgcgcc tgcccttttc ctggcgcttt ctgtgcgct 180
gttttagtgc cataaagtag aatacttgcc actagaaccg gagacattac gccatgaaca 240
agagcgccgc cgctggcctg cctggcctag cccgcgtcag caccgacgac caggacttga 300
ccaaccaacg ggccgaactg cagcggcccg gctgcaccaa gctgttttcc gagaagatca 360
ccggcaccag gcgcgaccgc ccggagctgg ccaggatgct tgaccaccta cgccctggcg 420
acgtttgtac agtgaccagg ctagacogcc tggcccgcag caccgcgac ctactggaca 480
ttgccgagcg catccaggag gccggcgccg gcctgcgtag cctggcagag ccgtgggccc 540
acaccaccac gccggccggc cgcattggtg cgcacagat cgcgggcatt gccgagttcg 600
agcgttccct aatcatcgac cgcaccogga gcgggcgcga ggccgccaag gcccgaggcg 660
tgaagttagg cccccgccct accctcacc cgcacagat cggcacagat gcgcacgccc 720
tcgaccagga aggcgcgacc gtgaaagagg cggctgcact gcttggcgtg catcgctcga 780
ccctgtaccg cgcacttgag cgcagcgagg aagtgcgccc caccgaggcc aggcggcgcg 840
gtgccttccg tgaggacgca ttgaccgagg cccagccctt ggccggccgc gagaatgaac 900
gccaagagga acaagcatga aaccgcacca ggacggccag gacgaaccgt ttttcattac 960
cgaagagatc gaggcggaga tgatcgcgcc cgggtacgtg ttogagccgc ccgcgcacgt 1020
ctcaaccgtg cggctgcatg aaatcctggc cggtttgtct gatgccaaagc tggcgccctg 1080
gccggccagc ttggccgctg aagaaaacga gcgcgcgctg ctaaaaaggt gatgtgtatt 1140
tgagtaaaac agcttgcgtc atgcggtcgc tgcgtatatg atgcgatgag taaataaaca 1200
aatacgcaag gggaaacgcat gaaggttatc gctgtactta accagaaagg cgggtcaggc 1260
aagacgacca tcgcaaccca tctagccccc gccctgcaac tcgccggggc cgtgttctg 1320
ttagtgcatt ccgatcccca gggcagtgcc cgcgattggg cggccgtgcg ggaagatcaa 1380
ccgctaaccg ttgtcggcat cgaccgccc cagattgacc ggcagctgaa ggccatcgcc 1440
cggcgcgact tcgtagtgat cgacggagcg cccaggcgg gtgcagccaa tgtgtccgcg 1500
atcaaggcag ccgacttcgt gctgattccg gtgcagccaa gcccttacga catatgggccc 1560
accgcccacc tgggtggagc gggttaagcag cgcattgagg tcacggatgg aaggctacaa 1620
gcggcccttg tcgtgtcgcg ggcgatcaaa ggcacgcgca tcggcggtga ggttgccgag 1680
gcgctggccg ggtacgagct gcccattctt gagtcccgtg tcacgcagcg cgtgagctac 1740
ccaggcactg ccgcccgcgg cacaaccggt cttgaatcag aacccgaggg cgacgctgcc 1800
cgcgaggtcc aggcgctggc cgtgaaatt aaatcaaaac tcatttgagt taatgaggta 1860
aagagaaaaa gagcaaaagc acaaacacgc ccgtccgagc ccgtccgagc gcacgcagca 1920
gcaaggctgc aacgttggcc agcctggcag acacgccagc catgaagcgg gtcaactttc 1980
agttgcccgg ggaggatcac accaagctga agctaccaga ggtacgcca ggaagagca 2040
ttaccgagct gctatctgaa tacatcgccc agctaccaga gtaaatgagc aaatgaataa 2100
atgagttagt ggctaaagga ggcggcatgg ggcggcatgg gtaaatgagc acaaccaggc 2160
accgacgccc tggaatgccc catgtgtgga ggaacgggag gttggccagg cgtaaaggcg 2220
tgggttgtct gcggccctg caatggcact ggaaccccca agcccagagg atcgcgctga 2280
cggtcgcaaa ccatccggcc cggtagaaat cggcgcgcg gaggcagaag acctggttga 2340
gaagttagag gccgcgcagg ccgcccagcg gcaacgcata ccccggaac cagccccggg 2400
tgaatcgtgg caagcggccc ctgatcgaat ccgcaagaa tcccggcaac cgccggcagc 2460
```

cggtgcgcgcg	tcgattagga	agccgccc	gggagacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tccgagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagaggtt	tccgcagggc	cggccggcat	ggccagtggt	tgggattacg	acctggtact	2700
gatggcgggt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacgggat	ccgaggggtg	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggcccg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccggtttc	tctaccgcct	3120
ggcacgccgc	gcccgcaggc	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggaggt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atgtgaagga	ggaggcgggg	caggctggcc	cgatccctagt	3300
catgcgctac	cgcgaacctg	tcgagggcga	agcatccgcc	ggttccta	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaagggtcga	aaagggtctc	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	actttaaaaa	actcttaaaa	cccgcctggc	3600
ctgtctgata	ctgtctggcc	agcgacacgc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgccttc	gcgtcggcct	atcgccggcc	ctggccgcctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accaggggcg	ggacaagccg	cgccgtcgcc	3780
actcgaccac	cggcgccac	atcaaggcac	ctcgctcgc	gcgtttcggt	gatgacggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggtgccc	3900
ggagcagaca	agcccgtcag	ggcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagcca	3960
tgaccagctc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggt	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcac	aggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggctcgctc	4140
gctgcggcga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataaacga	ggaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtg	ctggcggttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttcccc	4380
tggaagctcc	ctcgctgcgt	ctcctgttcc	gacctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgcttcc	tactagctca	cgctgtagg	atctcagttc	4500
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccggtc	agcccgaccg	4560
ctgcgcctta	tcgggtaact	atcgctctga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actgacagca	gccattggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	ccagttacct	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	agcgggtggt	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacacaa	4800
caccgctggg	gatcctttga	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaaag	4860
atctcaagaa	agcgggtggt	tctttcttac	ggggtctgac	gctcagtgga	acgaaaaactc	4920
acgttaaggg	atcttcttga	tgcatcttag	gtactaaaac	aattcatcca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatatcct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgcctcg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccagg	cgccgtggga	aaagacaagt	tcctcttcgg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgcgcggatc	tttaaattgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tcggccagat	cggtattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctatcc	gtatagggac	aatccgat	gtcgatggag	tgaagagacc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggct	ttgttcactc	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcctcactca	tgagcagatt	gtccagacca	tcattgcccgt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgctattttt	aaatatagg	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cttctcttcc	gtatctttta	cgcagcggtg	5700
tttttcgac	agtttttttc	attccgggtg	tattctcatt	ttagccattt	attatttctc	5760
tctcttttcc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcaactgt	tccttgcatt	ctaaaacctt	aaataaccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaagt	tgccgtataa	catagtatcg	acggagccga	ttttgaaaac	gcgggtgatca	5940
caggcagcaa	cgtctgtgca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6060
tctgcccgc	tacaacggct	ctcccgtgta	cgccgtcccg	gactgatggg	ctgcctgtat	6120
cgagtggtga	ttttgtgccc	agctgcccgt	cggggagctg	ttggctggct	gggtggcagga	6180
tatattgtgg	tgtaaaacaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taattgtactg	aattaacgcc	gaattaatcc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acatactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcata	tctcggtgac	gggcaggacc	6480

ggacggggcg	gtaccggcag	gctgaagtc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgcttcgtgc	6600
atggcgacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccgt	ccgctgggtg	6720
cggggggaga	cgtacacggt	cgactcggcc	gtccagtcgt	aggcgttgcg	tgcttccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgc	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgctg	ctcggtagcg	6900
aagttgaccg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgacagccgc	cggcagtgtc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgctcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatat	agaggaaagt	cttgcgaaag	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	tttgggacca	ctgtcggcag	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcttttat	cgcaatgatg	gcatttgtag	gtgccacctt	ccttttctac	7320
tgtccttttg	atgaagtgc	agatagctgg	gcaatggaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggtcttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtagggt	ccaccttctc	ctttctactgt	ccttttgatg	aagtgcagag	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gtccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	cgcctctctc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	cgttatcacag	acatgataag	atacattgat	8100
gagtttgagc	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaaatttgt	8160
gatgctattg	ctttatttgt	aaccattata	agctgcaata	aacaagttgg	ggtggggcaa	8220
gaactccagc	atgagatccc	cgcgctggag	gatcatccag	ccggcgtccc	ggaaaacgat	8280
tccgaagccc	aaactttcat	agaagggcgc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgctcc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgcccccac	ggctgctcgc	cgatctcggt	catggccggc	ccggaggcgt	cccggaagtt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctggtc	ctggaccgcg	ctgatgaaca	gggtcacgct	8580
gtcccgagcc	acacgggcga	agtgcctctc	cgcgaagtcc	cgggagaacc	cgagccgggtc	8640
ggtccagaac	tcgacgcctc	cggcgacgct	gcgcgcgggtg	agcaccggaa	cggcactgggt	8700
caacttggcc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaatcc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaaaggg	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaaggaaggt	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaggcta	tcgttcaaga	tgctctcgcc	gacagtggtc	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggctat	tgagactttt	caacaaaggg	9180
taatatcggt	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaaggtggc	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtgggtcca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttctc	ctatataagg	9480
aagttcattt	catttggaga	ggacacgctg	aaatcaccag	tctctctcta	caaactctatc	9540
tctctcgagc	tttcgcagat	ccgggggggc	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggaggcgtg	ggatatgtcc	9720
tgccggtaaa	tagctgcgcc	gatgggttct	acaaagatcg	ttatgtttat	cggcactttg	9780
catcggccgc	gctcccgaat	ccggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgtg	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgtgtg	tctacaaccc	gtcgcggagg	ctatggatgc	gatcgctgcg	gccgatctta	9960
gccagacgag	cgggttcggc	ccattcggac	cgcaaggaa	cggtaataac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggagc	10080
acacgctcag	tgcgtccgtc	gcgagggtc	tcgatgagct	gatgcttttg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcggtc	attgactgga	gcgaggcgat	gttcggggat	tcccaatacg	10260
aggtcgccaa	catcttcttc	tggaggccgt	ggttggtctg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	gagcttgacg	actccggcgc	tatatgctcc	tatatgctcc	10380
gcattggtct	tgaccaactc	tatcagagct	tggttgacgg	gatgcagctt	gatgcagctt	10440
gggcgcaggg	tcgatgcgac	gcaatcgctc	gatccggagc	gggcgtacac		10500

aaatcgcccg	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaacccg	acgccccagc	actcgctccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattagg	ttcctatagg	gtttcgctca	tgtgttaggc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	accaaatacc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcggtta	ttcagatcaa	gcttgacctg	10860
gaatatcgcg	agtaaaactga	aaatcacgga	aaatgagaaa	tacacacttt	aggacgtgaa	10920
atatggcgag	gaaaactgaa	aaaggtggaa	aatttagaaa	tgtccactgt	aggacgtgga	10980
atatggcaag	aaaactgaaa	atcatggaaa	atgagaaaaca	tccacttgac	gacttgaaaa	11040
atgacgaaat	cactaaaaaa	cgtgaaaaat	gagaaatgca	cactgaagga	ctccgcggga	11100
attcgattgt	gctagccaat	gtttaacaag	atgtcaagca	caatgaatgt	tgggtggttg	11160
tgggtcggtg	tggcggtggt	ggaaaattgc	gggtggttcga	gcggtagtga	tcggcgatgg	11220
ttgggtgttg	cagcggtgtt	tgatatcgga	atcacttatg	gtgggttgta	caatggaggt	11280
gcgtcatggt	tattgggtggt	tggtcaccta	tatatatttt	taataatatt	aagtattttta	11340
cctatttttt	acatatatttt	tattaaattt	atgcattggt	tgtatttttt	aatagttttt	11400
atcgctactg	ttttataaaa	tattttatta	ttttatgtgt	tatatattta	cttgatgtat	11460
tggaaaatttt	ttccattggt	ttttctatat	ttataataat	tttcttattt	ttttttgttt	11520
tattatgtat	tttttcgttt	tataataaat	atttattaaa	aaaaatatta	tttttgtaaa	11580
atatatcatt	tacaatgttt	aaaagtcatt	tgtgaatata	ttagctaagt	tgtacttctt	11640
tttgtgcatt	tgatgtgtgt	catgtctatt	atgttctctt	ggccaaaaca	tgtctactcc	11700
tgtcacttgg	gttttttttt	ttaagacata	atcactagt	attatatcta	gactgaaggc	11760
gggaaacgac	aatctgatca	tgagcggaga	atgaagggag	tcacgttatg	accccccgcc	11820
atgacgcggg	acaagccgtt	ttacgtttgg	aactgacaga	accgcaacgt	tgaaggagcc	11880
actcagccgc	gggtttctgg	agtttaatga	gctaagcaca	tacgtcagaa	accattattg	11940
cgcgttcaaaa	agtcgcctaa	ggtcactatc	agctagcaaa	tatttcttgt	caaaaatgct	12000
ccactgacgt	ttcataaatt	ccccctcggt	tccaattaga	gtctcatatt	cactctcaat	12060
ccaaataatc	tgacacggat	ctcgagatcg	aattccccgc	gcccgcgaatt	cactagtggg	12120
tccccgggta	cggtcagtc	cttatgttac	gtcctgtaga	aaacccaacc	cgtgaaatca	12180
aaaaactcga	cggcctgtgg	gcattcagtc	tggatcgcca	aaactgtgga	attgagcagc	12240
ggtggtggga	aagcgcgtta	caagaaaagcc	gggcaattgc	tgtgccaggc	agtttttaacg	12300
atcagttcgc	cgatgcagat	attcgtaatt	atgtgggcaa	cgtctggtat	cagcgcgaag	12360
tctttataacc	gaaagggttg	gcaggccagc	gtatcgtgct	gcgtttcgat	gcggtcactc	12420
attacggcaa	agtgtgggtc	aataatcagg	aagtgatgga	gcacagggc	ggctatacgc	12480
cattttgaagc	cgatgtcacg	ccgtatgtta	ttgccgggaa	aagtgtacgt	atcacagttt	12540
gtgtgaacaa	cgaactgaac	tggcagacta	tcccgccggg	aatggtgatt	accgacgaaa	12600
acggcaagaa	aaagcagctc	tacttccatg	atttctttaa	ctacgccggg	atccatcgca	12660
gcgtaatgct	ctacaccacg	ccgaacacct	gggtggacga	tatcacccgtg	gtgacgcatg	12720
tcgcgcaaga	ctgtaaccac	gcgtctgttg	actggcaggt	ggtggccaat	ggtgatgtca	12780
gcgttgaact	gcgtgatgcg	gatcaacagg	tgggtgcaac	tggacaaggc	accagcggga	12840
cttttgcaagt	ggtgaatccg	cacctctggc	aaacgggtga	aggttatctc	tatgaactgt	12900
acgtcacagc	caaaagccag	acagagtgtg	atatctaccc	gctgcgcgtc	ggcatccggt	12960
cagtggcagt	gaagggcgaa	cagttccctga	tcaaccacaa	accgtttctac	tttactggct	13020
ttggccgtca	tgaagatgcg	gatttgcgcg	gcaaaaggatt	cgataacgtg	ctgatggtgc	13080
acgatcacgc	attaatggac	tggattgggg	ccaactccta	ccgtacctcg	cattaccctt	13140
acgtgaaga	gatgtctgac	tgggcagatg	aacatggcat	cgtggtgatt	gatgaaactg	13200
cagctgtcgg	ctttaacctc	tctttaaggca	ttggtttcga	agcgggcaac	aagccgaaaag	13260
aactgtacag	cgaagaggca	gtcaacgggg	aaactcagca	ggcgcaactta	caggcgatta	13320
aagagctgat	agcgcgtgac	aaaaaccacc	caagcgtggt	gatgtggagt	attgccaacg	13380
aacgggatac	ccgtccgcaa	ggtgcacggg	aatatttctgc	gccactggcg	gaagcaacgc	13440
gtaaactcga	tccgacgcgt	ccgatcacct	gcgtcaatgt	aatgtttctgc	gacgctcaca	13500
ccgataccat	cagcgatctc	tttgatgtgc	tgtgcctgaa	ccgttattac	ggttggtatg	13560
tccaaagcgg	cgattttgaa	acggcagaga	aggtactgga	aaaagaactt	ctggcctggc	13620
aggagaaact	gcatcagccg	attatcatca	ccgaatacgg	cgtgggatac	ttagccgggc	13680
tgcactcaat	gtacaccgac	atgtggagt	aagagtatca	gtgtgcatgg	ctggatatgt	13740
atcacccgct	ctttgatcgc	gtcagcgcgc	tcgtcggtga	acaggtatgg	aatttcgccc	13800
attttgcgac	ctcgcaaggc	atattgcgcg	ttggcggtaa	caagaagggg	atcttcaccc	13860
gcgaccgcaa	accgaagtgc	gcggcttttc	tgtgcaaaaa	acgctggact	ggcatgaact	13920
tcggtgaaaa	accgcagcag	ggaggcaaac	aatgaatcaa	caactctcct	ggcgacccat	13980
cgtcggttac	agcctcgggg	attgcgtacc	gagctcgaat	ttccccgatc	gttcaaacat	14040
ttggcaataa	agttttcttaa	gattgaatcc	tgttgcgggt	cttgcgatga	ttatcatata	14100
atttctgttg	aattacgtta	agcatgtaat	aattaacatg	taatgcatga	cgttattttat	14160
gagatgggtt	tttatgatta	gagtcgccga	attatacatt	taatacgcga	tagaaaacaa	14220
aatatagcgc	gcaaactagg	ataaattatc	gcgcgcgggt	tcatctatgt	tactagatcg	14280
ggaattcgat	atcaagcttg	gcactggccg	tcgtttttaca	acgtcgtgac	tgggaaaacc	14340
ctggcgttac	ccaacttaat	cgccttgcag	cacatcccc	tttcgcccagc	tggcgtaata	14400
gcgaagaggc	ccgcaccgat	cgccttccc	aacagttgcg	cagcctgaat	ggcgaatgct	14460
agagcagctt	gagcttggat	cagattgtcg	tttccgcct	tcagtttaaa	ctatcagtgt	14520

ttgacaggat	atattggcgg	gtaaacctaa	gagaaaagag	cgttttattag	aataacggat	14580
attttaaaagg	gcgtgaaaaag	gtttatccgt	tcgtccattt	gtatgtg		14627

<210> 110
 <211> 9080
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18attBZeo(6XHS4) 2eGFP Plasmid

<400> 110						
cagttgcccgg	ccggggtcggc	cagggcggaac	tcccgccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccggaa	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccg	caccacacac	caggccagg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtcccga	ccacaccggc	gaagtcgtcc	240
tcacgaagt	cccgggagaa	cccagcccg	tcggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatata	gaattcctgc	420
agcccccg	atccgctcac	ggggacagcc	ccccccaaa	gccccagg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	gtccggcgct	540
ccccccgcat	ccccgagccg	gcagcgtcgg	gggacagccc	gggcacgggg	aagggtggcac	600
ggggatcgctt	ttctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgctaggg	ggcagcagcg	agccgcccgg	ggctccgctc	cggtccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggcacgg	ggaaggtggc	840
acgggatcgc	acgggatcgc	acgcttctcg	agcctcttgc	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggac	agcccccccc	caaagcccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagccgccc	ggggctccgc	tccgggtccgg	1020
cgctcccccc	gcatccccga	gcccggcagc	tgccggggaca	gcccggggcac	ggggaagggtg	1080
gcacgggatc	gctttcctct	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcggtatcc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcgagccgc	ccgggggctcc	gctccgggtcc	1260
ggcgctcccc	ccgcatcccc	gagccggcag	cgctgcggga	cagccccggg	acgggggaagg	1320
tggcacggga	tcgctttcct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggccgcggat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440
gtaattacgt	ccctcccccg	ctagggggca	gcagcgagcc	gcccggggct	ccgctccggt	1500
ccggcgctcc	ccccgcatec	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
ggtggcacgg	gatcgctttc	ctctgaacgc	ttctcgctgc	tttttgagcc	tgcagacacc	1620
tgggggatac	ggggccgcgg	atccgctcac	ggggacagcc	ccccccaaa	gccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aagggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaataaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgccctg	1980
gctgaccgcc	caacgacccc	cgcccattga	cgtaataaat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgccact	2100
tggcagtaga	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	atttagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcggggcga	ggcgggaggg	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagt	ccttttatgg	cgaggcggcg	gcggcgcgcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cggtgccttc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccgcgc	cggtcttgac	tgaccgcgtt	actcccacag	gtgagcgggg	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgcttgg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaaag	cttaaagggg	tccgggaggg	ccctttgtgc	ggggggggag	ggctcggggg	2760
gtgctgctgc	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	cccgcgctgc	ccggcggtcg	2820
tgagcgtcgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccgggggg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaagg	tgcggtcggg	2940
gtgtgtgcgt	gggggggtga	gcagggggtg	tgggcgcggc	ggctcgggct	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggtt	cggtgcgggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgccgggc	gggggggtgg	ggcaggtggg	ggtgccgggg	3120
ggggcggggg	cgctccgggc	cgggggaggg	tcgggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcggtcg	tcgaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcagg	acttctcttg	tcccaaatct	ggcggagccg	aaatctggga	ggcgccggcg	3300

cacccccctct	agcggggcgcg	ggcggaagcgg	tgcgggcgcg	gcaggaagga	aatggggcggg	3360
gagggccttc	gtgcgctcgc	gcgcgcgcgt	cccccctctcc	atctccagcc	tcggggctgc	3420
cgagggggga	cggctgcctt	cggggggggac	ggggcagggc	ggggttcggc	ttctggcgtg	3480
tgaccggcgg	ctctagagcc	tctgtctaacc	atgttcatgc	cttcttcttt	ttcctacagc	3540
tccctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	tggcaaagaa	ttcgccacca	3600
tggtgagcaa	gggcgaggag	ctgttcaccg	gggtgggtgcc	catcctggtc	gagctggacg	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	tggcccaccc	3780
tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccccagc	cacatgaagc	3840
agcacgactt	cttcaagtc	gccatgccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagacccgcg	ccgaggtgaa	gttcgagggc	gacaccctgg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcgtg	cagctcgccg	4140
accactacca	gcagaacacc	cccatcgccg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	gaccccaacg	gaagcgcgat	cacatggtcc	4260
tgctggagtt	cgtgaccgcc	gccgggatca	ctctcgccat	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaaggtggt	ggctggtgtg	gccaatgccc	4380
tggtcacaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccctga	gcactcgact	tctggcta	aaaggaaatt	tattttcatt	gcaatagtgt	4500
gttggaattt	tttgtgtctc	tcactcggaa	ggacatatgg	gagggcaaat	cattttaaac	4560
atcagaatga	gtatttggtt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaaggtggc	tataaagagg	tcatacgtat	atgaaacagc	cccctgctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttctttta	catccctaaa	attttccctta	catgtttttac	tagccagatt	tttctctctc	4800
tccctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	gcacctgcag	4860
cccaagcttg	catgctcgca	ggtcgactct	agtggatccc	ccgcccgtta	tccccaggtt	4920
gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	tgccaccctc	4980
cccgctgccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgccggggga	gcgccggacc	5040
ggagcggagc	cccgggcgcc	tcgctgctgc	tgccctagcg	gggagggagc	taattacatc	5100
cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	cgcggccccc	tatccccag	5160
gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	aaagcgatcc	cgtgccacct	5220
tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	gatgcggggg	gagcgccgga	5280
cgggagcggg	gccccgggcg	gctcgtgctg	gccccctagc	gggggagggg	cgtaattaca	5340
tccctggggg	ctttgggggg	gggctgtccc	cgtgagcggg	tccgcggccc	cgatatcccc	5400
aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	cccgtgccac	5460
cttccccgtg	cccggtgtgt	cccgccacgc	ccctcggtcg	gggatgcggg	gggagcgccg	5520
gaccggagcg	gagccccggg	cggctcgtcg	ctgcccccta	gcgggggagg	gacgtaatta	5580
catccctggg	ggctttgggg	gggggctgtc	cccgtagcgc	gatccgcggc	cccgtatccc	5640
ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	atcccgctgc	5700
accttccccg	gtctgcaggc	gtccccgcac	gctggcggct	cggggatgcg	gggggagcgc	5760
cggaccggag	cggagccccg	ggcggtcgcg	tgctgcccc	tagcggggga	gggacgta	5820
tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	cggtatccgc	gccccgtatc	5880
ccccaggtgt	ctgcagctc	aaagagcagc	gagaagcgtt	cagaggaaag	cgatcccggt	5940
ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgcggg	ctcggggatg	cggggggagc	6000
gccgggacgg	agcggagccc	cgggcgggct	gctgctgccc	cctagcgggg	gagggagcga	6060
attacatccc	tgggggcttt	gggggggggg	tgtccccgtg	agcggatccg	cggccccgta	6120
tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	6180
tgccaccctc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgccggggga	6240
gcgcgggacc	ggagcggagc	cccgggcggc	tcgctgctgc	cccctagcgg	gggagggagc	6300
taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	cgcggggctg	6360
caggaattcg	taatcatggt	catagctgtt	tctgtgtgta	aattgttatc	cgctcacaat	6420
tccacacaac	atagcagccg	gaagcataaa	gtgtaaagcc	tgggtgctct	aatgagtgag	6480
ctaactcaca	ttaattgcgt	tgcgctcact	gcccgccttt	cagtcgggaa	acctgtcgtg	6540
ccagctgcat	taatgaatcg	gccaacgcgc	ggggagaggc	ggtttgcgta	ttggggcgctc	6600
ttccgcttcc	tcgctcactg	actcgtcgcg	ctcggctcgt	cggctgcggc	gagcgggtatc	6660
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataacg	caggaaagaa	6720
catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	6780
tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	cagcgtcaa	gtcagagggtg	6840
gcgaaacccg	acaggactat	aaagatacca	ggcggtttccc	cctggaagct	ccctcgtgcg	6900
ctctcctggt	ccgacctgc	cgcttacccg	ataacctgtc	gcctttctcc	cttcgggaa	6960
cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	7020
caagctgggc	tgtgtgcacg	aacccccctg	tcagcccgac	cgctgcgcct	tatccggtaa	7080
ctatcgtctt	gagtcacaac	cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	7140
taacaggatt	agcagagcga	ggtatgtagg	cgggtgctaca	gagttcttga	agtggtggcc	7200
taactacggc	tacactagaa	ggacagtatt	tggtatctgc	gctctgctga	agccagttac	7260
cttcggaaaa	agagttggta	gctcttgatc	cggcaaacaa	accacgcgtg	gtagcgggtg	7320

tttttttgggt	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	7380
gatcttttct	acggggtctg	acgctcagtg	gaacgaaaaa	tcacgttaag	ggatttttgg	7440
catgagatta	tcaaaaagga	tcttcaccta	gatcctttta	aattaaaaat	gaagttttta	7500
atcaatctaa	agtatatatg	agtaaacttg	gtctgacagt	taccaatgct	taatcagtg	7560
ggcacctatc	tcagcgatct	gtctatttcg	ttcatccata	gttgccctgac	tccccgctg	7620
gtagataact	acgatacggg	agggcttacc	atctggcccc	agtgtctgcaa	tgataccgcg	7680
agaccacagc	tcaccggctc	cagattttatc	agcaataaac	cagccagccg	gaagggccga	7740
gcgcagaagt	ggctcctgcaa	ctttatccgc	ctccatccag	tctattaatt	gttgccggga	7800
agctagagta	agtagttcgc	cagttaatag	tttgcgcaac	gttggttgcca	ttgctacagg	7860
catcgtgggtg	tcacgctcgt	cgtttggtat	ggcttcattc	agctccggtt	cccaacgatc	7920
aaggcgagtt	acatgatccc	ccatgtttgtg	caaaaaagcg	gttagctcct	tcgggtcctcc	7980
gatcgtttgtc	agaagtaagt	tggccgcagt	gttatcactc	atgggttatgg	cagcactgca	8040
taattctctt	actgtcatgc	catccgtaag	atgcttttct	gtgactgggtg	agtactcaac	8100
caagtcatto	tgagaatagt	gtatgcccgc	accgagttgc	tcttgcccgg	cgtcaataacg	8160
ggataataacc	gcgccacata	gcagaacttt	aaaagtgtct	atcattggaa	aacgttcttc	8220
ggggcgaaaa	ctctcaagga	tcttacgcgt	gttgagatcc	agttcagatgt	aaccactcgt	8280
tgacccaac	tgatcttcag	catcttttac	tttcaccagc	gtttctgggt	gagcaaaaac	8340
aggaaggcaa	aatgccgcaa	aaaagggaat	aaggcgacac	cggaaatgtt	gaatactcat	8400
actcttcctt	tttcaatatt	attgaagcat	ttatcagggt	tattgtctca	tgagcggata	8460
catatttgaa	tgtatttaga	aaaataaaca	aatagggggt	ccgcgcacat	ttccccgaaa	8520
agtccacact	gacgtagtta	acaaaaaaaa	gcccgccgaa	gcgggcttta	ttaccaagcg	8580
aagcgccatt	cgccattcag	gctgcgcaac	tgttggaag	ggcgatcgggt	gcgggctct	8640
tcgctattac	gccagctggc	gaaaggggga	tgctgtgcaa	ggcgattaaag	ttgggtaacg	8700
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtccgtaata	cgactcactt	8760
aaggccctga	ctagagggtc	gacggtatac	agacatgata	agatacattg	atgagtttgg	8820
acaaaccaca	actagaatgc	agtgaaaaaa	atgctttatt	tgtgaaattt	gtgatgctat	8880
tgctttattt	gtaaccatta	taagctgcaa	taaacaagtt	gggggtggcg	gaagaactcca	8940
gcatgagatc	cccgcgctgg	aggatcatcc	agccggcgtc	ccggaaaacg	attccgaagc	9000
ccaacctttc	atagaaggcg	gcggtggaat	cgaatctcg	tagcacgtgt	cagtcctgct	9060
cctcgccac	gaagtgcacg					9080

<210> 111

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA10 Plasmid

<400> 111

gttaactacg	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	60
tttctaata	cattcaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttatccctt	180
ttttgcccga	ttttgccttc	ctgtttttgc	tcaccagaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggtg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagtctt	360
gctatgtggc	gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgcgcat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tgcatgaca	gtaagagaat	tatgcagtc	tgccataacc	atgagtgata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaaacat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgctgtagc	aatggcaaca	acgttgcgca	aactattaac	720
tggggaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	780
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	840
tgagccgggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	900
ctcccgatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	ggatagagtg	cctcactgat	taagcattgg	taactgtcag	accaagttaa	1020
ctcatatata	cttttagattg	atttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttggt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aaatcaaaaag	aatagcccga	gatagggttg	agtgttggtc	cagtttgga	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggcgaaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggt	cgaggtgccg	taaagcacta	1380
aatcggaacc	ctaaaaggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgctgcgcgt	aaccaccaca	cccgccgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctaggtga	agatcctttt	tgataatctc	atgacaaaaa	tcccttaacg	tgagttttcg	1620


```
ttccactgag cgtcagaccc cgtagaaaaag atcaaaggat cttottgaga tccttttttt 1680
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttggttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gottcagcag agcgcagata 1800
ccaaatactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgcctacat acctcgctct gctaactctg ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 1980
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagagaaa ggcggacagg 2100
tatccggtaa gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac 2160
gcctggtatc tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg 2220
tgatgctcgt cagggggggc gaggcctatg aaaaacgcga gcaacgcggc ctttttacgg 2280
ttcctggcct tttgctggcc ttttgctcac atgtaatgtg agttagctca ctcattaggc 2340
accccaggct ttacacttta tgcctccggc tgcctatgtg tgtggaattg tgagcggata 2400
acaatttcac acaggaaaca gctatgacca tgattacgcc aagctacgta atacgactca 2460
ctagtggggc ccgtgcaatt gaagccgggt ggcgccaagc ttctctgcag gattgaagcc 2520
tgctttttta tactaacttg agcgaaatct ggatcaccat gaaaacattt aacatttctc 2580
aacaagatct agaattagta gaagtagcga cagagaagat tacaatgctt tatgaggata 2640
ataaacatca tgtgggagcg gcaattcgta cgaaaacagg agaaatcatt tcggcagtag 2700
atattgaagc gtatatagga cgagtaactg ttgtgcaga agccattgcy attggtagt 2760
cagtttcgaa tggacaaaag gattttgaca cgattgtagc tgtagacac ccttattctg 2820
acgaagtaga tagaagtatt cgagtggtaa gtccttgtgg tatgtgtagg gagttgattt 2880
cagactatgc accagattgt tttgtgttaa tagaaatgaa tggcaagtta gtcaaaaacta 2940
cgattgaaga actcattcca ctcaaatata ccgaaaatta aaagttttac cataccaagc 3000
ttggctgctg cctgaggctg gacgacctcg cggagtctta ccggcagtgcc aaatccgtcg 3060
gcatccagga aaccagcagc ggctatccgc gcatccatgc ccccgaaactg caggagtggg 3120
gaggcacgat ggccgctttg gtccggatct ttgtgaagga accttacttc tgtggtgtga 3180
cataattgga caaactacct acagagattt aaagctctaa aaatttttaa 3240
gtgtataatg tgttaaaacta ctgattctaa ttgtttgtgt attttagatt 3300
gaactgatga atgggagcag tgggtggaatg cctttaatga ggaaaacctg ttttgctcag 3360
aagaaaatgcc atctagtgat gatgaggcta ctgctgactc tcaacattct actcctccaa 3420
aaaagaagag aaaggtagaa gaccccaagg actttccctc agaattgcta agttttttga 3480
gtcatgctgt gtttagtaat agaactcttg cttgctttgc tatttacacc acaagggaaa 3540
aagctgcact gctatacaag aaaattatgg aaaaatattc tgtaaccttt ataagtaggc 3600
ataacagtta taatcataac atactgtttt ttcttactcc acacaggcat agagtgtctg 3660
ctattaataa ctatgctcaa aaattgtgta ccttttagctt tttaatgtt agaggggtta 3720
ataaggaata tttgatgtat agtgccctga ctagagatca taatcagcca taccacattt 3780
gtagaggttt tacttgcttt aaaaaacctc ccacacctcc cctgaacct gaaacataaa 3840
atgaatgcaa ttgttgttgt taacttgttt attgcagctt ataatgggtt caaataaagc 3900
aatagcatca caaatttcac aaataaagat ccacgaattc gctagcttcg gccgtgacgc 3960
gtctccggat gtacaggcat gcgtcgaccc tctagcaag gccttaagt agtcgtatta 4020
cggactggcc tcgtgtttac aacgtcgtga ctgggaaaac cctggcggtt cccaacttaa 4080
tcgccttgca gcacatcccc cttcgcccag ctggcgtaat agcgaagagg cccgcaccga 4140
tcgccttccc caacagttgc gcagcctgaa tggcgaaatg cgcttcgctt ggtaataaag 4200
ccgccttcgg cgggcttttt ttt 4223
```

<210> 112
<211> 5855
<212> DNA
<213> Artificial Sequence

<220>
<223> pCX-LamIntR Plasmid

```
<400> 112
gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgccg ggctgacccg 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggactttcca ttgacgtcaa tgggtgggact atttacggta aactgcccac ttggcagtag 240
atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtag atgaccttat gggactttcc tacttggcag tacatctacg 360
tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
gcgatggggg cggggggggg gggggcgcgc gccagggggg gcggggcggg gcgagggggc 540
gggcggggcg agggcgagag gtgcgggcgg gcctaataaa aagcgaagcg cgcggcgggc 600
tccttttatg gcgaggcggc cgccccgtgc cccgctccgc gccgcctcgc gccgcccggc 720
gggagtcgct ccggttcgct tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780
```

gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaa	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggctcgggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gccgcgcgtg	cccgccggct	gtgagcgctg	960
cgggcgcggc	gcggggcttt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccggggggc	1020
ggtgccccgc	ggtgcggggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	ggtgtgtgcg	1080
tggggggggtg	agcagggggg	gtggggcgcg	cggctcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggccccggct	tcgggtgcgg	ggctccgtgc	ggggcggtggc	1200
gcgggggctcg	ccgtgccggg	cgggggggtgg	cggcaggtgg	gggtgccggg	cggggcgggg	1260
ccgectcggg	ccgggggagg	ctcgggggag	gggcgcggcg	gccccggagc	gcgggcggct	1320
gtcagggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgcgagg	1380
gacttccttt	gtcccaaato	tggcggagcc	gaaatctggg	aggcgccggc	gcacccccctc	1440
tagcggggcgc	ggcggaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggccctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccctctc	catctccagc	ctcggggctg	ccgcagggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cgggggttcgg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgtctaac	catgttcattg	ccttctctctt	tttctacag	ctcctgggca	1680
acgtgctggt	tgttgtgctg	tctcatcatt	ttggcaaaga	attcatggga	agaaggcgaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaggacgggt	aaagagtttg	gattaggcag	agacaggcga	atcgcaatca	1860
ccgactcggg	acaggccaac	attgagttat	tttcaggaca	caaacacaag	cctctgacaa	1920
cgagaatcaa	cagtataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tcctggccag	cagaggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgtccac	ttgaagacat	caccacaaaa	gaaattgcgg	2100
caatgctcaa	tggatcata	gacgagggca	aggcgcgctc	agccaagtta	atcagatcaa	2160
cactgagcga	tgcattccga	gaggcaatag	ctgaaggcca	tataacaaca	aaccatgtcg	2220
ctgccactcg	cgcagcaaaa	tctagagtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaaaatttta	tcaagcagca	gaatcatcac	catgtttggc	cagacttgca	atggaactgg	2340
ctgttggttac	cgggcaacga	gttggtgatt	tatgcgaaat	gaagtggctc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattggccatc	ccaacagcat	2460
tgcataattga	tgtctcggga	atatcaatga	aggaaacact	tgataaatgc	aaagagattc	2520
tgggcgagga	aaccataatt	gcactactc	gtcgcaacc	gctttcatcc	ggcacagtat	2580
caagggtat	tatgcgcgca	cgaaaagcat	caggctcttcc	cttcgaaggg	gatccgccta	2640
cctttcacga	gttgccgag	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgtcaaca	tcttctcggg	cataagtccg	acaccatggc	atcacagtat	cgtgatgaca	2760
gaggcagggga	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820
cctatcagaa	ggtggtggct	ggtgtggcca	atgccctggc	tcacaaatac	cactgagatc	2880
tttttccctc	tgccaaaaat	tatggggaca	tcgtgaagcc	ccttgagcat	ctgactctcg	2940
gctaataaag	gaaattttatt	ttcattgcaa	tagtgtgttg	gaattttttg	tgtctctcac	3000
tcggaaggac	atatgggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttggtttaga	3060
gtttggaac	atatgccata	tgctggctgc	catgaacaaa	ggtggctata	aagagggtca	3120
tagtcaaca	aacagccccc	tgctgtccat	tccttattcc	atagaaaagc	cttgacttga	3180
ggttagat	tttttatatt	ttgttttgtg	ttattttttt	ctttaacatc	cctaaaattt	3240
tccttacatg	ttttactagc	cagatttttc	ctcctctcct	gactactccc	agtcatagct	3300
gtccctcttc	tcttatgaag	atccctcgac	ctgcagccca	agcttggcgt	aatcatggct	3360
atagctgttt	cctgtgtgaa	attgttatcc	gtcacacatt	ccacacacaa	tacgagccgg	3420
aagcataaag	tgtaaagcct	ggggtgccta	atgagttagc	taactcacat	taattgcgtt	3480
gcgctcactg	cccgccttcc	agtcgggaaa	cctgtcgtgc	cagcggatcc	gcactcfaat	3540
tagtcagcaa	ccatagctcc	gcccctaact	ccgcccattc	cgcccctaac	tcgcgccagt	3600
tcgcgccatt	ctccgcccc	tggctgacta	atttttttta	tttatgcaga	ggccgaggcc	3660
gcctcggcct	ctgagctatt	ccagaagtag	tgaggaggct	tttttggagg	cctaggcttt	3720
tgcaaaaagc	taacttggtt	attgcagctt	ataatggtta	caaataaagc	aatagcatca	3780
caaatttcac	aaataaagca	tttttttcac	tgcattctag	ttgtgggttg	tccaaactca	3840
tcaatgtatc	ttatcatgtc	tggatccgct	gcattaatga	atcgcccaac	gcgcggggag	3900
aggcggtttg	cgtattgggc	gtctctccgc	ctctcgcctc	actgactcgc	tgcgctcggg	3960
cgttcggctg	tatcagctca	tatcagctca	ttcctaaggc	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	4080
taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	4140
aaatcgagca	tcaagtcaga	ggtggcgaaa	ccgcagagga	ctataaagat	accaggcggt	4200
tccccctggc	agctccctcg	tgcgctctcc	tggtccgacc	ctgcccgtta	ccggatacct	4260
gtccgccttt	ctcccttcgg	gaagcggtgg	gctttctcaa	tgctcacgct	gtaggtatct	4320
cagttcgggtg	taggtcggtc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gcttatccg	gtaactatcg	tcttgagtc	aacccggtaa	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	4500
tacagagttc	ttgaagtggt	ggcctaacta	cggctacact	agaaggacag	tatttggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	4620
acaaaccacc	gctggtagcg	gtgggttttt	tgtttgcaag	cagcagatta	cgcgacagaa	4680
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	4740
aaactcacgt	taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	4800

ttttaatttaa	aaatgaagtt	ttaaatacaat	ctaaagtata	tatgagtaaa	cttgggtctga	4860
cagttaccaaa	tgcttaataca	gtgaggcacc	tatctcagcg	atctgtctat	ttcggttcac	4920
catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	4980
ccccagtgct	gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	5040
aaaccagcca	gccggaaggg	cgcgagcgag	aagtggctct	gcaactttat	ccgcctccat	5100
ccagttctatt	aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	5160
caacgtttgt	gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	5220
attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	5280
agcggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	5340
actcatgggt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	5400
ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	5460
ttgctcttgc	ccggcgtaaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	5520
gctcatcatt	ggaaaaagtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	5580
atccagttcg	atgtaaacca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	5640
cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaaatgcc	gcaaaaaagg	gaataagggc	5700
gacacggaaa	tgttgaatac	tcataactctt	cctttttcaa	tattattgaa	gcattttaca	5760
gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	5820
ggttccgcgc	acatttcccc	gaaaagtgcc	acctg			5855

<210> 113

<211> 4346

<212> DNA

<213> Artificial Sequence

<220>

<223> pSV40-193AttpsensePur Plasmid

<400> 113

ccggtgccgc	caccatcccc	tgaccacgc	ccctgacccc	tcacaaggag	acgaccttcc	60
atgaccgagt	acaagccac	gggtgcgcctc	gccacccgcg	acgacgtccc	ccgggcccga	120
cgcaccctcg	ccgcgcggtt	cgccgactac	cccgcacgc	gccacaccgt	cgaccccggac	180
cgccacatcg	agcgggtcac	cgagctgcaa	gaactcttcc	tcacgcgcgt	cgggctcgac	240
atcggcaagg	tgtgggtcgc	ggacgacggc	gcccgcggtg	cggtctggac	cacgccggag	300
agcgtcgaag	cgggggcggt	gttcgcccgc	atcggcccgc	gcatggccga	ggtgagcggg	360
tcccggctgg	ccgcgcagca	acagatggaa	ggcctcctgg	cgccgcaccg	gcccgaaggag	420
cccgcgtggt	tcctggccac	cgtcggcgtc	tcgcccagcc	accagggcaa	gggtctgggc	480
agcgcgcgtc	tgctccccgg	agtggaggcg	gcccgcgcgc	ccgggggtgc	cgcttctctg	540
gagacctccg	cgccccgcaa	cctccccctc	tacgagcggc	tcggcttcac	cgtcaccgcc	600
gacgtcgagg	tgcccgaagg	accgcgcacc	tggtgcatga	cccgcgaagg	cggtgcctga	660
cgcccgcccc	acgaccgcga	gcgcccgcac	gaaaaggagcg	cacgacccca	tggctccgac	720
cgaagccgac	ccgggcggcc	ccgcccagcc	cgcacccgcc	cccgaaggcc	accgactcta	780
gaggatcata	atcagccata	ccacatttgt	agagggtttta	cttgctttta	aaaacctccc	840
acacctcccc	ctgaacctga	aacataaaat	gaatgcaatt	gttggtgtta	acttggttat	900
tgcagcttat	aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	960
tttttctactg	cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	atcatgtctg	1020
gatccgcgcc	ggatccttaa	ttaagtctag	agtcgactgt	ttaaacctgc	aggcatgcaa	1080
gcttggcgta	atcatgggtca	tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacaaattc	1140
cacacaacat	acgagccgga	agcataaagt	gtaaagcctg	gggtgcctaa	tgagtgagct	1200
aactcacatt	aattgcgttg	cgctcactgc	ccgctttcca	gtcgggaaac	ctgtcgtgcc	1260
agctgcatta	atgaatcggc	caacgcgcgg	ggagaggcgg	tttgcgattt	gggcgctctt	1320
ccgcttcctc	gctcactgac	tcgctgcgct	cggtcggtcg	gctgcggcga	gcggtatcag	1380
ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	1440
tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcggtg	ctggcggtttt	1500
tccataggtg	ccgccccctc	gacgagcatc	acaaaaaatcg	acgctcaagt	cagaggtggc	1560
gaaacccgac	aggactataa	agataccagg	cgtttccccc	tggaaagctcc	ctcgtgcgct	1620
ctcctgttcc	gaccctgccc	cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	1680
tggcgctttc	tcatagctca	cgtgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	1740
agctgggctg	tgtgcacgaa	ccccccgttc	agcccagacc	ctgcgcctta	tccggtaact	1800
atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	1860
acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tggtggccta	1920
actacggcta	cactagaagg	acagtatttg	tctgtctgag	tctgtctgaag	ccagttacct	1980
tcggaaaaag	agttggtagc	tcttgatccg	gcaaaacaaac	caccgctggg	agcgggtgggt	2040
tttttgtttg	caagcagcag	attacgcgca	gaaaaaaaagg	atctcaagaa	gatcctttga	2100
tcttttctac	ggggtctgac	gctcagtgga	acgtaaaactc	acgttaaggg	attttggtca	2160
tgagattatc	taaaaaggatc	ttcacctaga	tccttttaaaa	taaaaaatga	agtttttaaat	2220
caatctaaag	tatatatgag	taaacttggg	ctgacagtta	ccaatgctta	atcagtgagg	2280
cacctatctc	agcgatctgt	ctatttcggt	catccatagt	tgccctgactc	cccgtcgtgt	2340

```

agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 2400
acccacgctc accggctcca gatttatcag caataaacca gccagccgga agggcccgagc 2460
gcagaagtgg tccctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 2520
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tggtgccatt gctacaggca 2580
tcgtgggtgc acgctcgtcg tttgggtatgg cttcattcag ctccgggtcc caacgatcaa 2640
ggcgagttac atgatcccc atgttggtgca aaaaagcggg tagctccttc ggtcctccga 2700
tcgttggtcag aagtaagttg gccgcagtggt tatcactcat gggtatggca gcactgcata 2760
attctcttac tgtcatgcca tccgtaagat gcttttctgt gactgggtgag tactcaacca 2820
agtcattctg agaatagtggt atgcggcgac agagtgtgctc ttgcccggcg tcaatacggg 2880
ataataccgc gccacatagc agaacttttaa aagtgtcat cattggaaaa cgttcttcgg 2940
ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 3000
caccacaactg atcttcagca tcttttactt tcaccagcgt tctctgggtga gcaaaaacag 3060
gaaggcaaaa tggccgcaaaa aaggggaataa gggcgacacg gaaatggtga atactcatac 3120
tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 3180
tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 3240
tgccacctga cgtctaagaa accattatta gtgatgacgg tgaaaacctc tgacacatgc 3300
tcacgaggcc ctttcgtctc gcgcgtttcg gctgtgtctgt aagcggatgc cgggagcaga caagcccgtc 3360
agctcccggg gacggtcaca gctgtgtctgt ggggctgggt taactatgcg gcacagatgc gcatcagagc 3420
agggcgcgct agcgggtgtt ggccgggtgtc ggggctgggt gcacagatgc gtaaggagaa 3480
agattgtact gagagtgcac catatgcggg tgcgaatacc ggctgcgcaa ctgttgggaa 3540
aataccgcat caggcgccat tccgcatcca cgccagctcg cgaaaggggg atgtgctgca aggcgattaa 3600
tgccgggctc ttcgctatta tccagtcag caggttgtaa aacgacggcc aggtcctccc gcaggcagaa 3660
gttgggtaac gccagggttt tttaggtgtg aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3720
agctgtggaa tgtgtgtcag ttaggggtgt caactagtc tcaattagtc agcaaccata gtcccgcgcc 3780
gtatgcaaa gcatgcctc catctagcaa aagtagtcaa ctaactccgc ccatctccg ccccatggct 3840
cagcaggcag catcccgc ttttatttat aggtttttt catctaagta gttttttatg tttacagtat 3900
taactccgc gactaattt aggtttttt gttttttatg tttacagtat atcattttac 4020
agtagtgagg aggtttttt gttttttatg tttacagtat atcattttac 4080
tcactaatac gttttttatg tttacagtat atcattttac 4140
tatgtagtct gttttctcgt tttacagtat atcattttac 4200
gtttctcgtt tttacagtat atcattttac 4260
tggtgcaacg cccactccct gctctctggg ggccgcg 4320
cccactccct gctctctggg ggccgcg 4346

```

<210> 114
 <211> 3166
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18attBZeo Plasmid

```

<400> 114
cagttgccgg ccgggtcgcg cagggcgaac tcccgcccc acggctgctc gccgatctcg 60
gtcatggccg gccgggaggg gtcccggaa ttcgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacac caggccaggg tggtgtccgg caccacctgg 180
tccctggacc cgctgatgaa cagggtcacg tccgtcccga ccacaccggg gaagtgcgtc 240
tccacgaagt cccgggagaa cccgagccgg tccgtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgcatgc ctgcaggtcg 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaata agtgagctaa ctcacattaa ttgcgttgcg ctactgccc 600
gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg 660
agaggcgggt tgcgtattgg gcgctcttcc gcttccctgc tactgactc gctgcgctcg 720
gtcgttcggc tgccgcgagc ggtatcagct cactcaaaagg cggtaatacg gttatccaca 780
gaatcagggg cgaataaagg tgagcaaaag gccagcaaaa ggccaggaac 840
cgtaaaaagg ccgcgttgct ggcggttttc cataggctcc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtca gaggtggcga aaccgcagac gactataaag ataccaggcg 960
tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgcccgt taccggatac 1020
ctgtccgctc tttcccttc cgtgtttctc atagctcacg ctgtaggtat 1080
ctcagttcgg tgtaggtcgt tccgtccaag ctgggctgtg tgcacgaacc ccccgttcag 1140
cccgaccgct gcgccttata cagtgttaac cgtcttgagt ccaaccgggt aagacacgac 1200
ttatcgccac tggcagcagc aggttagca aggttagca gagcagggt tgtaggcggt 1260
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

```

```

aaacaaacca cgcgtggttag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tcttttgatc ttttctacgg ggtctgacgc tcagtggaaac 1500
gaaaactcac gttaaaggat ttttggtcatg agattatcaa aaaggatctt cacttagatc 1560
cttttaaat aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 1620
gacagttacc aatgcttaat cagttaggca cctatctcag cgatctgtct atttcgttca 1680
tccatagttag cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaaccagc cagccggaag ggccgagcgc agaatgggtc ctgcaacttt atccgcctcc 1860
atccagtgta ttaattgttg ccgggaagct cccgggagct gtggtgtcac gctcgtcgtt taatagtttg 1920
cgcaacggtt ttgccattgc tacaggcatc cgagttacat gatcccccat gttgtgcaaa 2040
tcattcagct ccggttccca tctctccgac tctcttactg tcatgccatc gtaagtggc cgcagtgtta 2100
aaagcgggta gctccttcgg actgcataat ctcaaccaag tctcttactg tcagtccatc cgtaagatgc 2160
tcactcatgg ttatggcagc ctcaaccaag aataccgcgc ccaactgat cttcagcatc ggggaataagg 2220
ttttctgtga ctgggtgagta aataccgcgc cgaaaactct caaggatctt accgctgttg 2280
agtgtctctt gcccggcgtc ttcttcgggg cactcgtgca cccaactgat cttcagcatc ttttactttc 2340
gtgctcatca ttggaaaaac agatccagtt ctgggtgagc aaaaacagga actcatactc cgcgaaaaaa 2400
agatccagtt cgatgtaacc accagcgttt aatggtgaat cgggatacata ggcgaaaaatg aatattattg 2460
accagcgttt ctgggtgagc aaaaacagga actcatactc ttcttttttc tttgaataaa taacacaaata 2520
cgacacggga aatggtgaat cgggatacata ccacgtgacg tagttaacaa aaaaaagccc 2580
cagggttatt gtctcatgag gcacatttcc caagcgaagc gccattcgcc attcaggctg cgcaactggt 2640
ggggttccgc gctttattac atcggtgcgg gctcttcgac ggttttccca gtcacgacgt attcaggctg 2700
gccgaacggg gctttattac atcggtgcgg gctcttcgac ggttttccca gtcacgacgt attcaggctg 2760
gggaaggggc attaagttgg gtaacgccac tcaacttaagg ccttgactag agggctcgacg gtatatacag 2820
ctgcaaggcg gtaatacgac acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc 2880
atgataagat tttatttggg aaatttggta tgcattttaa ccattataag ctgcaataaa 2940
caagtggggg tgggcgaaga actccagcat gagatccccg cgctggagga tcatccagcc 3000
ggcgtcccg gaaacgattc cgaagcccaa cctttcatag aaggcggcgg tggaatcgaa 3060
atctcgtagc acgtgtcagt cctgctcctc ggccacgaag tgcacg 3120

```

<210> 115

<211> 7600

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo3'6XHS4eGFP Plasmid

<400> 115

```

cagttgccc cggggtcgcg cagggcgaac tcccgcctcc acggctgctc gccgatctcg 60
gtcatggccg gccgggaggg gtcccgaag ttcgtggaca cgacctccga ccactcgggc 120
tacagctcgt ccaggccgcg caccacacac caggccaggg tgttggtccg caccactggg 180
tccgtggaga cgtgtatgaa cagggtcacg tctgtccgga ccacaccggc gaagtgcgtc 240
tccacgaagt cccgggagaa cccgagccgg tccgtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatcta gttattaata 420
gtaatacaat acgggggtcat tagttcatag cccatataag gagttccgag ttacataact 480
tacggtaaat ggccgcctcg gctgaccgcc caacgacccc cgccatttga cgtcaataat 540
gacgtatggt cccatagtaa cgccaatagg gactttccat tgacgtcaat ggggtgacta 600
tttacggtaa actgcccact tggcagtaca tcaagtgtat catatgccaa gtacgcccc 660
tattgacgtc aatgacggta aatggcccg cttggtcatt gcccagtaga tgacctatg 720
ggactttcct acttggcagt acatctacgt attagtcatc gctattacca tgggtcgagg 780
tgagccccc gttctgcttc actctcccc tctccccccc ctccccaccc ccaattttgt 840
atttatttat tttttaatta ttttgtgcag cgatgggggc gggggggggg gggggcgcg 900
ccaggcgggg cggggcgggg cgagggcgcg ggccggggcg ggcggagagg tgccggcgga 960
gccaatcaga gcggcgcgct ccgaaaagtt ccttttatgg cgaggcgggc gcggcgggcg 1020
ccctataaaa agcgaagcgc cgggcggggc ggagtcgctg cgttgccctc gccccgtgcc 1080
ccgctccgcg ccgctcgcg cggcccgccc cggctctgac tgaccgcgtt actcccacag 1140
gtgagcgggc gggacggccc ttctcctcgg ggtgtgaatt agcgtttggt ttaatgacgg 1200
ctcgtttctt ttctgtggct gcgtgaaagc cctaaagggc tccgggaggg cctttgtgc 1260
gggggggagc ggctcggggg gtgctgctgc gtgtgtgtgc gtggggagcg ccgctgctgc 1320
cccgcgctgc ccggcggtc tgagcgtgct gggcgcgggc gggggctttg tgcgtccgc 1380
gtgtgcgcga ggggagcgcg gccggggggc gtgccccgcg gtgcgggggg gctgcgagg 1440
gaacaaaggg tgcgtgcggg gtgtgtgcgt ggggggggtg gcaggggggt tgggcgcggc 1500
ggtcgggctg taaccccccc ctgcaccccc ccccccgagt tgcgtgagcac ggcccggtt 1560
cgggtgcggg gtcctgtgcg gggcgtggcg cggggctcgc cgtgccgggc ggggggtggc 1620

```

ggcaggtggg	ggtgcccggg	ggggcggggc	cgcctcgggc	cggggagggc	tccggggagg	1680
ggcgccggcg	ccccggagcg	ccggcgcgctg	tcgagggcg	gcgagccgca	gccattgcct	1740
tttatggtaa	tcgtgcgaga	ggggcgaggg	acttcctttg	tcccaaactc	ggcggagccg	1800
aaatctggga	ggcgccgccc	cacccccctct	agcggggcg	ggcgaagcgg	tggggcgccg	1860
gcaggaagga	aatgggagg	gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	1920
atctccagcc	tcggggcgctg	cgcaggggga	cggctgcctt	cgggggggac	ggggcagggc	1980
gggggttcggc	ttctggcgctg	tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	2040
cttcttcttt	ttcctacagc	tcctgggcaa	cgtgctgggt	ggtgtgctgt	ctcatcattt	2100
tggcaaagaa	ttcgccacca	tggtgagcaa	ggcgaggag	ctgttcaccg	gggtgggtgc	2160
catcctgggtc	gagctggacg	gcgacgtaaa	ggcccaaacg	ttcagcgtgt	ccggcgaggg	2220
cgagggcgat	gccacctacg	gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	2280
gcccgtgccc	tgggccaccc	tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	2340
ctaccccgac	cacatgaagc	agcacgactt	cttcaagtcc	gccatgcacc	aaggctacgt	2400
ccaggagcgc	accatcttct	tcaaggacga	cggaactac	aagaccgcg	ccgaggtgaa	2460
gttcgagggc	gacaccctgg	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	2520
cggaacatc	ctggggcaca	agctggagta	caactacaac	agccacaacg	tctatatcat	2580
ggccgacaag	cagaagaacg	gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	2640
cggcagcggtg	cagctcgccg	accactacca	gcagaacacc	cccatcgccg	acggccccgt	2700
gctgctgccc	cacaaccact	acctgagcac	ccagtccgcc	ctgagcaaac	accccaacga	2760
gaagcgcgat	gacatggctc	tgctggagtt	cgtgaccgcc	gcccggatca	ctctcgccat	2820
ggacgagctg	tacaagtaag	aattcactcc	tcaggtgcag	gctgcctatc	agaagggtgt	2880
ggctgggtgtg	gccaatgccc	tggttcacaa	atacactga	gatctttttc	cctctgcca	2940
aaattatggg	gacatcatga	agccccttga	gcatctgact	tctggctaatt	aaaggaaatt	3000
tattttcatt	gcaatagtgt	gttggaaatt	tttgtgtctc	tactcggaa	ggacatatgg	3060
gagggcaaat	cattttaaacc	atcagaatga	gtatttggtt	tagagtgttg	caacatatgc	3120
catatgctgg	ctgccatgaa	caaagggtgg	tataaagagg	tcatacagat	atgaaacagc	3180
cccctgctgt	ccatttcctta	ttccatagaa	aagccttgac	ttgagggttag	atttttttta	3240
tattttgttt	tgtgttattt	ttttctttta	catccctaaa	attttccctta	catgttttac	3300
tagccagatt	tttccctcctc	tcctgactac	tcccagtcac	agctgtccct	cttctcttat	3360
gaagatccct	cgacctgcag	cccaagcttg	ggtcgactct	ggtcgactct	agtggtatccc	3420
ccgccccgta	tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	3480
agcgatcccc	tgccaccttc	cccgtgcocg	ggctgtcccc	gcacgctgcc	ggctcgggga	3540
tgccgggggga	gcgcgggacc	ggagcggagc	cccgggcggc	tcgctgctgc	cccctagcgg	3600
gggagggacg	taattacatc	cctgggggct	ttgggggggg	gctgtcccg	tgagcggatc	3660
cgcgggcccc	tatccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	3720
aaagcgatcc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	3780
gatgcggggg	gagcgccgga	ccggagcggg	gccccggg	gctcgctgct	gccccctagg	3840
gggggaggga	cgtaattaca	tccttggggg	ctttgggggg	gggctgtccc	cgtgagcggg	3900
tccgcggccc	cgtatccccc	aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	3960
ggaaagcgat	cccgtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgccggctcg	4020
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggtcgctgt	ctgcccccta	4080
gcgggggagg	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	4140
gatccgcggc	cccgtatccc	ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	4200
gaggaaagcg	atcccgtgcc	accttccccg	tgccccggct	gtccccgcac	gctgcggcgt	4260
cggggatgcg	ggggggagcg	cggaccggag	cggagccccg	ggcggtctgc	tgctgcccc	4320
tagcggggga	gggacgtaat	tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	4380
cggatccgcg	gccccgtatc	ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	4440
cagaggaaag	cgatcccgtg	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgccgg	4500
ctcggggatg	cggggggagc	gccggaccgg	agcggagccc	cgggcggtct	gctgctgcc	4560
cctagcgggg	gaggggagta	attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	4620
agcggatccg	cggccccgta	tccttcaggc	gtctgcaggc	tcaaagagca	gcgagaagcg	4680
ttcagaggaa	agcgtatccc	tgccaccttc	cccgtgcccg	ggctgtcccc	gcacgctgcc	4740
ggctcgggga	tgccggggga	gcgcgggacc	ggagcggagc	cccggggcgg	tcgctgctgc	4800
cccctagcgg	gggagggagc	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	4860
tgagcggatc	cgcggggctg	caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	4920
aattgtttatc	cgctcacaat	tccacacaa	atacagagcc	gaagcataaa	gtgtaaagcc	4980
tggggtgcct	aatgagtgag	ctaactcaca	ttaattgcgt	tgcgctcact	gcccgttttc	5040
cagtcgggaa	acctgtcgct	ccagctgcat	taatgaatcg	gccaacgcgc	gggggagagg	5100
ggtttgcgta	ttgggcgctc	ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcggt	5160
cggctgcggc	gagcgggtatc	agctcactca	aaggcggtaa	tacggttatc	cacagaatca	5220
ggggataacg	caggaaagaa	catgtgagca	aaaaggccag	ctgacgagca	gaaccgtaaa	5280
aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	aaagataacca	tcacaaaaat	5340
cgacgctcaa	gtcagagggtg	gcgaaaccgg	acaggactat	cgcttaccgg	ggcggtttcc	5400
cctggaagct	ccctcgctgcg	ctctcctggt	ccgacctgc	cagctgtgag	gtatctcagt	5460
gcctttctcc	cttcgggaag	cgtggcgctt	tcctatagct	aacccccgt	tcagccccgac	5520
tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	cggtgaagaca	cgacttatcg	5580
cgtgcgcct	tatccggtaa	ctatcgtctt	gagtcacaac			5640

```
ccactggcag cagccactgg taacaggatt agcagagoga ggtatgtagg cgggtgctaca 5700
gagttccttga agtgggtggcc taactacggc tacactagaa ggacagtatt tggatatctgc 5760
gctctgctga agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa 5820
accaccgctg gtagcggtgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa 5880
ggatctcaag aagatccttt gatcttttct acgggggtctg acgctcagtg gaacgaaaac 5940
tcacgttaag ggatttttgg catgagatta tcaaaaagga tcttcaccta gatcctttta 6000
aattaaaaaat gaagttttta atcaatctaa agtatatatg agtaaaacttg gtctgacagt 6060
taccaatgct taatcagtg ggcacctatc tcagcgatct gtctatttcg ttcattccata 6120
gttgccctgac tccccgcgtg gtagataaact acgatacggg agggccttacc atctggcccc 6180
agtgcctgcaa tgataccgcg agaccacgac tcaccggctc cagatttatc agcaataaac 6240
cagccagccg gaagggccga ggcgcagaagt ggtcctgcaa ctttatccgc ctccatccag 6300
tctattaatt gttgcccggg catcgtgggtg acatgatccc agtaagtaag tggccgcagt 6360
gttgcttgcca ttgctacagg aaggcgagtt gatcgttgct agaaagtaag catccgtaag 6420
agctccgggtt cccaacgatc aagcgagtt gatcgttgct agaaagtaag catccgtaag 6480
gttagctcct tcggctcctc taattctctt tgagaaatag gtatgcggcg accgagttgc 6540
atgggttatgg cagcactgca caagtcattc tgagaatag gtatgcggcg accgagttgc 6600
gtgactgggtg agtactcaac cgtcaataac ggataatacc gcgccacata gcagaacttt 6660
tcttgcccgg cgtcaataac ggataatacc gcgccacata gcagaacttt 6720
atcattggaa aacggttctt tgcacccaac tcttcaagga tcttaccgct gttgagatcc 6780
agtccgatgt agtactcaac cgtcaataac ggataatacc gcgccacata gcagaacttt 6840
gtttctgggt gagcaaaaaac aggaaggcaa aatgccgcaa aaaagggaat aaggggcgaca 6900
cggaatggt gaatactcat actcttctct tttcaatatt attgaagcat ttatcagggt 6960
tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggtt 7020
ccgcgcacat ttccccgaaa agtgccacct gagctagtta acaaaaaaaa gcccgccgaa 7080
gcgggcttta ttaccaagcg aagcgccatt cgccattcag gctgcgcaac tgttgggaag 7140
ggcgatcggt gcgggcctct tcgctattac gccagctggc gaaaggggga tgtgctgcaa 7200
ggcgattaaag ttgggtaacg ccagggtttt cccagtcacg acgttgtaaa acgacggcca 7260
gtccgtaata cgactcactt aaggccttga ctagaggggtc gacggtatac agacatgata 7320
agatacattg atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt 7380
tgtgaaattt gtgatgctat tgccttattt gtaaccatta taagctgcaa taaacaagtt 7440
gggggtggcg aagaactcca gcatgagatc cccgcgctgg aggatcatcc agccggcgctc 7500
ccggaanaac attccgaagc ccaacctttc atagaaggcg gcggtggaat cgaaatctcg 7560
tagcacgtgt cagtccctgct cctcggccac gaagtgcacg 7600
```

<210> 116

<211> 7631

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo5'6XHS4eGFP Plasmid

<400> 116

```
cagttgcccg cggggtcgcg cagggcgaac tccccccccc acggctgctc gccgatctcg 60
gtcatggccg gcccggaggc gtcccgaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacac cagggccaggg tggttgtccg caccactgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagccgg tcgggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgcg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatata gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc ccccccacaa gccccagggt atgtaattac 480
gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg gtccggcgct 540
ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aagggtggcac 600
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
acggggcccg ggatccgctc acggggacag ccccccccca aagccccagc ggatgtaatt 720
acgtccctcc cccgctaggg ggcagcagcg agccgcccgg ggtccgctc cggtccggcg 780
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc 840
acgggatcgc tttcctctga acgcttctcg ctgctctttg agcctgcaga cacctggggg 900
atacgggggc gcggatccgc tcacggggac agcccccccc caaagcccc caggatgtaa 960
ttacgtccct ccccgctag ggggcagcag cgagccgccc ggggtccgc tccgggtccg 1020
cgctcccccc gcatccccga gccggcagcg tgccggggaca ggggaagggt ggggaagggt 1080
gcacgggatc gctttctct gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
ggatacgggg ccgcgatcc gctcacgggg acagcccccc cccaaagccc ccagggatgt 1200
aattacgtcc ctcccccgct aggggggacg agcgagccgc ccgggggtcc gctccggtcc 1260
ggcgctcccc ccgcatcccc gagccggcag cgtgcgggga cagccccggc acgggggaag 1320
tggcacggga tcgctttcct ctgaacgctt ctcgctgctc tttgagcctg cagacacctg 1380
ggggatacgg ggcccgcgat ccgctcacgg ggacagcccc ccccaagac cccagggat 1440
```


gtaattacgt	ccctccccc	ctagggggca	gcagcgagcc	gccccgggct	ccgctccggt	1500
ccggcgctcc	ccccgcaccc	ccgagccggc	agcgtgcccc	gacagcccg	gcacggggaa	1560
ggtggcacgg	gatcgctttc	ctctgaacgc	ttctcgctgc	tctttgagcc	tgacagacacc	1620
tgggggatac	ggggccgcgg	atccgctcac	ggggacagcc	ccccccaaa	gccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgccccggg	ctccgctccg	1740
gtccggcgct	ccccccgcat	ccccgagccg	gcagcggtcg	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctcttttag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgccctg	1980
gctgaccgcc	caacgacccc	cgcccatgga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggaacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gccagttaca	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tggttcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atthatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggcggggg	2400
cgagggggcg	ggcgggggcg	ggcgggggcg	tgccggcgca	gccaatcaga	gcggcgcgct	2460
ccgaaagttt	cctttttatg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtcgctg	cggttcgctt	gccccgtgcc	ccgctccgcg	ccgctccgcg	2580
ccgcccgcct	cggtctctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgtttggt	ttaatgacgg	ctcgtttcct	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tcggggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgctgctgt	gtgtgtgtgc	gtggggagcg	cccgctgctg	cccgctgctg	ccggcggtcg	2820
tgagcgctgc	ggggcgggcg	cggttcgctt	tgctgcgag	gtgtgcgca	ggggagcgcg	2880
gccggggggc	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgctgcgggg	2940
gtgtgtgcgt	gggggggggt	gcaggggggt	tgggcgcgcg	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggtt	cggtgagggg	gctccgtgcg	3060
gggctgtggc	cggggctcgc	cgtgccgggc	gggggggtgg	ggcaggtggg	ggtgcggggc	3120
ggggcggggc	cgccctcgggc	cggggagggc	tcgggggag	ggcgcgcgcg	ccccggagcg	3180
ccggcggtcg	tcgagggcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
ggggcgaggg	acttcccttg	tcctcaaatc	ggcgagccg	aaatctggga	ggcgccgccc	3300
cacccccctc	agcggggcg	ggcgaaggcg	tgcgggcgcc	gcaggaagga	aatggggggg	3360
gagggccttc	gtgctgcgcc	gcgcgccgct	ccccctctcc	atctccagcc	tcgggggtcg	3420
cgccagggga	cggtctgcct	cgggggggac	ggggcagggc	gggggttcggc	ttctggcggtg	3480
tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	cttctctctt	ttcctacagc	3540
tcctgggcaa	cgtgctggtt	gttgtgctgt	ctcatcattt	tgccaaagaa	ttcgccacca	3600
tggtgagcaa	gggcgaggag	ctgttcaccg	gggttggtgc	catcctggtc	gagctggagc	3660
gcgacgtaaa	cgccacacaag	ttcagcgtgt	ccggcgaggg	cgaggcgcat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	tgggccaccc	3780
tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccocgac	cacatgaagc	3840
agcacgactt	cttcaagtcc	gccatgcccg	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcg	ccgaggtgaa	gttcgagggc	gacaccctgg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcataaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcggtg	cagctcgccg	4140
accactacca	gcagaacacc	cccatcgccg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	accccaacga	gaagcgcgat	cacatggctc	4260
tgctggagtt	cgtgaccgcc	gccgggatca	ctctcgccat	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaaggtggt	ggctggtgtg	gccaatgccc	4380
tggtcacaaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccttga	gcatctgact	tctggctaatt	aaaggaaatt	tattttcatt	gcaatagtgt	4500
gttgggaattt	tttgtgtctc	tcactcgga	ggacatatgg	gagggcaaat	catttaaaac	4560
atcagaatga	gtatttggtt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaaggtggc	tataaagagg	tcatacgtat	atgaaacagc	cccctgctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttctttta	catccctaaa	attttcctta	catgttttac	tagccagatt	tttcctcctc	4800
tccagactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	ggtcgactct	agaggatccc	agaggatccc	cgggtaccga	gctcgaattc	4920
gtaatcatgg	tcatagtgtg	ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	4980
catacgagcc	ggaagcataa	agtgtaaaag	ctgggggtg	taatgagtga	gctaactcac	5040
attaattgcg	ttgcgctcac	tgcccgcctt	ccagtcggga	aacctgtcgt	gccagctgca	5100
ttaatgaatc	ggccaacgcg	cggtttgcgt	cgggttgctg	attggcgctc	cttcgcttcc	5160
ctcgctcact	gactcgctgc	gctcggtcgt	tcggctgcgg	cgagcggtat	cagctcactc	5220
aaaggcggtg	atacggttat	ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	5280
aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	5340
gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	5400
gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	tcctcgtgc	gctctcctgt	5460

tccgaccctg	ccgcttaccg	gatacctgtc	cgccctttctc	ccttcgggaa	gcgtggcgct	5520
ttctcatagc	tcacgctgta	ggatctctag	ttcgggtgtag	gtcgttcgct	ccaagctggg	5580
ctgtgtgcac	gaacccccc	ttcagcccg	ccgctgcgc	ttatccggta	actatcgtct	5640
tgagtccaac	ccggttaagac	acgacttatc	gccactggca	gcagccactg	gtaacaggat	5700
tagcagagcg	aggatgtag	gcgggtgctac	agagttcttg	aagtgggtggc	ctaactacgg	5760
ctacactaga	aggacagtat	ttgggtatctg	cgctctgctg	aagccagtta	ccttcggaaa	5820
aagagttggg	agctcttgat	ccggcaaaaa	aaccacgcgt	ggtagcgggtg	gtttttttgt	5880
ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	gaagatcctt	tgatcttttc	5940
tacgggggtct	gacgctcagt	ggaacgaaaa	ctcacgttaa	gggatttttg	tcattgagatt	6000
atcaaaaagg	atcttcacct	agatcctttt	aaattaaaaa	tgaagtttta	aatcaatcta	6060
aagtatatat	gagtaaaactt	ggctctgacag	ttaccaatgc	ttaatcagtg	aggcacctat	6120
ctcagcgatc	tgtctatttc	gttcatccat	agttgcctga	ctccccgtcg	tgtagataac	6180
tacgatacgg	gagggcctac	catctggccc	cagtgctgca	atgataccgc	gagacccacg	6240
ctcacgggct	cagattttat	cagcaataaa	ccagccagcc	ggaagggccg	agcgcagaa	6300
tggtcctgca	actttatccg	cctccatcca	gtctattaat	tggtgcggg	aagctagagt	6360
aagtagttcg	ccagtttaata	gtttgcgcaa	cggtgttgcc	attgctacag	gcatcggtgt	6420
gtcacgctcg	tcggttggtg	tggttctatt	cagctccggt	tccaacgat	caaggcgagt	6480
tacatgatcc	cccattgtgt	gcaaaaaagc	ggttagctcc	ttcggctcctc	cgatcggtgt	6540
cagaagtaag	ttggccgcag	tggttatcact	catgggtatg	gcagcactgc	ataattctct	6600
tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggt	gagtactcaa	ccaagtcatt	6660
ctgagaatag	tgtatgcggc	gaccgagttg	ctcttgccc	gcgtcaatac	gggataatac	6720
cgccccacat	agcagaactt	taaaagtgct	catcattgga	aaacgttctt	cggggcgaaa	6780
actctcaagg	atcttaccgc	tggtgagatc	cagttcgatg	taaccactc	gtgcacccaa	6840
ctgatcttca	gcatctttta	ctttcaccag	cgtttctggg	tgagcaaaaa	caggaaggca	6900
aaatgccgca	aaaaagggaa	taagggcgac	acggaaatgt	tgaatactca	tactcttctt	6960
ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	atgagcggat	acataattga	7020
atgtatttag	aaaaataaac	aaataggggt	tccgcgcaca	tttccccgaa	aagtgccacc	7080
tgacgtagtt	aacaaaaaaa	agcccgccga	agcgggcttt	attaccaagc	gaagcgccat	7140
tcgccattca	ggctgcgcaa	ctgttgggaa	gggcgatcgg	tgccggccctc	ttcgctatta	7200
cgccagctgg	cgaagggggg	atgtgctgca	agggcattaa	gttgggtaac	gccagggttt	7260
tcccagtcac	gacgttgtaa	aacgacggcc	agtcctgaat	acgactcact	taaggccttg	7320
actagagggg	cgacggtata	cagacatgat	aagatacatt	gatgagtttg	gacaaaaccac	7380
aaactaattg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	ttgctttatt	7440
tgtaaccatt	ataagctgca	ataaacaagt	ttgggtgggc	gaagaactcc	agcatgagat	7500
ccccgcgctg	gaggatcatc	cagccggcgt	cccgaaaac	gattccgaag	cccaaccttt	7560
catagaaggc	ggcggtggaa	tcgaaatctc	gtagcacgtg	tcagtcctgc	tcctcggcca	7620
cgaagtgcac	g					7631

<210> 117

<211> 4615

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo6XHS4 Plasmid

<400> 117

cagttgccgg	ccgggtcgcg	cagggcgaa	ccccgcccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccg	caccacacc	caggccagg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtcccga	ccacaccggc	gaagtcgtcc	240
tcacgaaagt	cccgggagaa	cccgagccgg	tcgggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgcatgc	ctgcaggtcg	420
actctagtgg	atcccccgcc	ccgtatcccc	caggtgtctg	cagggtcaaa	gagcagcgag	480
aagcgttcag	aggaaagcga	tcccgtgcc	ccttccccgt	gccccgggctg	tccccgcacg	540
ctgccggctc	ggggatgccc	ggggagcgcc	ggaccggagc	ggagccccgg	gcggctcgt	600
gctgccccct	agcgggggag	ggacgtaat	acatccctgg	gggctttggg	ggggggctgt	660
ccccgtgagc	ggatccgccc	ccccgtatcc	cccaggtgtc	tgacggctca	aagagcagcg	720
agaagcgttc	agaggaaagc	gatccccgtg	caccttcccc	gtgcccgggc	tgtccccgca	780
cgctgccggc	tcggggatgc	ggggggagcg	ccggaccgga	gcggagcccc	ggcggtctcg	840
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	900
gtccccgtga	gcggatccgc	ggccccgtat	ccccaggtg	tctgcaggct	caaagagcag	960
cgagaagcgt	tcagaggaaa	gcgatcccgt	gccaccttcc	ccgtgcccgg	gctgtccccg	1020
caogctgccc	gctcggggat	gcggggggag	cgccggaccg	gagcggagcc	ccggggcggt	1080
cgctgctgcc	ccctagcggg	ggagggacgt	aattacatcc	ctgggggctt	tggggggggg	1140
ctgtccccgt	gagcggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	1200

```

agcgagaagc gttcagagga aagcgatccc gtgccacott ccccgtagcc gggctgtccc 1260
cgcacgctgc cggctcgggg atgccccggg agcgccggag cggagcggag ccccgggcgg 1320
ctcgctgctg cccctagcgc ggggagggac gtaattacat ccctgggggc tttggggggg 1380
ggctgtcccc gtgagcggat ccgcggcccc gtatccccca ggtgtctgca gggtcaaaga 1440
gcagcgagaa gcgttcagag gaaagcgatc ccgtgccacc ttccccgtgc ccgggtgtgc 1500
cccgcacgct gccggctcgg ggatgcgggg ggagcgccgg accggagcgg agccccgggc 1560
ggctcgtgct tgcctcctag cgggggaggg acgtaattac atccctgggg gctttggggg 1620
ggggctgtcc ccgtgagcgg atccgcggcc ccgtatcccc cagggtgtctg cagggtcaaa 1680
gagcagcgag aagcgttcag aggaaagcga tcccgtgcc ccttccccgt gcccgggctg 1740
tccccgcacg ctgcccgtct ggggatgcgg ggggagcgcc ggaccggagc ggagccccgg 1800
gcggctcgct gctgccccct agcgggggag ggacgtaatt acatccctgg gggctttggg 1860
ggggggctgt ccccgtaggc ggatccgcgg ggtgcagga attcgtaatc atggtcatag 1920
ctgtttcctg tgtgaaattg ttatccgctc tgccaatga gtgagctaac tcacattaat tgcgtttgcg 2040
ataaagtgtg aagcctgggg gggaaacctg gcgtattggg gggcgagcgc cttcctcgct cactgactcg 2100
tactgccccg ctttccagtc gcgcggggga gaggcggttt cggcgagcgc cttcctcgct cactgactcg 2160
cgcgcgggga gggcggttt cggcgagcgc gctgtcggct taacgcagga cgcgttgctg ggcgttttcc ataggctccg 2220
ctgctcctcg tgcgtcggct taacgcagga cgcgttgctg ggcgttttcc ataggctccg 2280
ttatccacag aatcagggga gtaaaaaagg ctcaagtcag aagctccctc tctoccttcg gtaggtcgtt 2340
gccaggaacc gtaaaaaagg ctcaagtcag aagctccctc tctoccttcg gtaggtcgtt 2400
gagcatcaca aaaaatcgag tccccctgg tgcgtcggct taacgcagga cgcgttgctg 2460
taccaggcgt tccccctgg tgcgtcggct taacgcagga cgcgttgctg 2520
accggatacc tgcgtcggct taacgcagga cgcgttgctg 2580
ttaggtatcc tgcgtcggct taacgcagga cgcgttgctg 2640
ccggttcagc cgcacccgctg tatcgccact ctacagagtt gtaggcgggtg 2700
agacacgact cgcacccgctg tatcgccact ctacagagtt gtaggcgggtg 2760
gtaggcgggtg gtaggcgggtg tatcgccact ctacagagtt gtaggcgggtg 2820
gtatttggtg tctgcgctct aacaaaccac aaaaaggatc aaaaactcac 2880
tgatccggca aaaaaggatc aaaaactcac aaaaatgaag atgcttaata 2940
acgcgcagaa aaaaaggatc aaaaactcac aaaaatgaag atgcttaata 3000
cagtggaacg aaaaaggatc aaaaactcac aaaaatgaag atgcttaata 3060
acctagatcc aaaaaggatc aaaaactcac aaaaatgaag atgcttaata 3120
acttggtctg acagttacca ccatagttgc ttaaccatca 3180
tttcgttcac ttaaccatca ttaaccatca ttaaccatca 3240
ttaccatca ttaaccatca ttaaccatca ttaaccatca 3300
ttaccatca ttaaccatca ttaaccatca ttaaccatca 3360
tccgcctcca tccagtcctat gcaacggtgt cgggttccaa 3420
aatagtttgc cgttcagctc aattcagctc cgttcagctc 3480
gggtatggct tttctgtgac gttgtcgttg tgggttccaa 3540
ttgtgcaaaa aagcgggttag cactcatggt tttctgtgac 3600
gcaagtgtat tttctgtgac gttgtcgttg tgggttccaa 3660
gtaagatgct gttgtcgttg gttgtcgttg tgggttccaa 3720
cggcgaccga tgcctcatcat ccagcgtttc cgacacggaa 3780
actttaaaag gatccagttc ccagcgtttc cgacacggaa 3840
ccgctgttga gatccagttc ccagcgtttc cgacacggaa 3900
tttactttca ccagcgtttc ccagcgtttc cgacacggaa 3960
ggaataaggg aggggttatt gggttccgcg cacatttccc 4020
agcatttatc gggttccgcg ccgaagcggg ctttattacc 4080
aaacaaatag gggttccgcg ccgaagcggg ctttattacc 4140
aaaaagcccc gcaactgttg ggaagggcga tgcgaaggcg 4200
gcaactgttg ggaagggcga tgcgaaggcg tgcgaaggcg 4260
ggggatgtgc gtaaaacgac ggccagtcgg taatacgact 4320
tatacagaca tgataagata cattgatgag tttggacaaa 4380
aaaaaatgct ttatttgtga aatttgtgat gctattgctt 4440
tgcaataaac aagttggggg gggcgaagaa ctccagcatg 4500
catccagccg gcgtcccggg aaacgattcc gaagcccaac 4560
ggaatcgaaa tctcgtagca cgtgtcagtc ctgctcctcg gccacgaagt 4615

```

<210> 118
 <211> 17384
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pFK161 Plasmid

<400> 118

gcgcacgagg	gagcttccag	ggggaaacgc	ctggatatctt	tatagtccctg	tcgggggtttc	60
gccacctctg	actttgagcgt	cgattttttgt	gatgctcgtc	agggggggcgg	agcctatgga	120
aaaacgcgcg	caacgcggcc	tttttacgggt	tcctggcctt	ttgctggcct	tttgctcaca	180
tggtctttcc	tcggttatcc	cctgattctg	tgataaacgc	tattaccgcc	tttgagtggag	240
ctgataccgc	tcgcccgcagc	cgaacgaccg	agcgcagcga	gtcagtgagc	gaggaagcgg	300
aagagcgctg	acttccgcgt	ttccagactt	tacgaaacac	ggaaaccgaa	gaccattcat	360
gttgttgctc	aggctcgaga	cgttttgcag	cagcagtcgc	ttcacgttcg	ctcgcgtatc	420
ggtgattcat	tctgctaacc	agtaaggcaa	ccccgccagc	ctagccgggt	cctcaacgac	480
aggagcacga	tcattgcgcac	ccgtcagatc	cagacatgat	aagatacatt	gatgagtttg	540
gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	600
ttgctttatt	tgtaaccatt	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	660
attttatggt	tcagggttcag	ggggagggtgt	gggagggtttt	ttaaagcaag	taaaacctct	720
acaaatgtgg	tatggctgat	tatgatctct	agtcaggcca	ctatacatca	aatattcctt	780
attaacccct	ttacaaatta	aaaagctaaa	ggtacacaat	ttttgagcat	agttattaat	840
agcagacact	ctatgcctgt	gtggagtaag	aaaaaacagt	atgttatgat	tataactggt	900
atgcctactt	ataaagggtta	cagaatattt	ttccataatt	ttcttgata	gcagtgcagc	960
tttttctctt	gtgggtgtaaa	tagcaaaagca	agcaagagtt	ctattactaa	acacagcatg	1020
actcaaaaaa	cttagcaatt	ctgaaggaaa	gtccttgggg	tccttctacct	ttctcttctt	1080
ttttggagga	gtagaatggt	gagagtcagc	agtagcctca	tcattcactag	atggcatttc	1140
ttctgagcaa	aacagggtttt	cctcattaaa	ggcattccac	cactgctccc	attcatcagt	1200
tccatagggt	ggaatctaaa	atacacaac	aattagaatc	agtagtttaa	cacattatac	1260
acttaaaaaa	tttatatttta	ccttagagct	ttaaatctct	gtaggtagtt	tgtccaatta	1320
tgctcacacca	cagaagtaag	gttccttcac	aaagatccgg	accaaagcgg	ccatcgctgc	1380
tccccactcc	tcaggttcgg	gggcattggat	gcgcggatag	ccgctgctgg	tttctgggat	1440
gccgacggat	ttgcaactgc	ggtagaactc	gcgaggtcgt	ccagcctcag	gcagcagctg	1500
aaccaactcg	cgaggggatc	gagccccggg	tgggcggaaga	actccagcat	gagatccccg	1560
cgctggagga	tcattccagcc	ggcgctcccg	aaaacgattc	cgaagcccaa	cctttcatag	1620
aaggcggcgg	tggaatcgaa	atctcgtgat	ggcaggttgg	gcgtcgcttg	gtcggtcatt	1680
tcgaacccca	gagtcgccgt	cagaagaact	cgtaagaag	gcgatagaag	gcgatgcgtt	1740
gcgaatcggg	agcggcgata	ccgtaaagca	cgaggaagcg	gtcagcccat	tcgcgcgcaa	1800
gctcttcagc	aatatcacgg	gtagcccaacg	ctatgtcctg	atagcgggtc	gccacacca	1860
gccggccaca	gtcgatgaat	ccagaaaagc	ggccattttc	caccatgata	ttcggcaagc	1920
aggcatcgcc	atgggtcacg	acgagatcct	cgccgtcggg	atgcgcgcct	tgagcctggc	1980
gaacagttcg	gctggcgaga	gcccctgatg	ctcttcgtcc	agatcatcct	gatcgacaag	2040
accggcttcc	atccgagtac	gtgctcgctc	gatgcgatgt	ttcgcttggt	ggtcgaatgg	2100
gcaggtagcc	ggatcaagcg	tatgcagccg	ccgcattgca	tcagccatga	tggtactttt	2160
ctcggcagga	gcaaggtgag	atgacaggag	atcctgcccc	ggcacttcgc	ccaatagctg	2220
ccagtccctt	cccgcttcag	tgacaacgtc	gagcacagct	gcgcaaggaa	cgcccgtcgt	2280
ggccagccac	gatagccgcg	ctgcctcgtc	ctgcagttca	ttcagggcac	cggacagggtc	2340
ggtcttgaca	aaaagaaccg	ggcgccccctg	cgctgacagc	cggaaacacg	cggcatcaga	2400
gcagcttgata	gtctgttggt	cccagtcata	gccgaatagc	ctctccacc	aagcggccgg	2460
agaacctgcg	tgcaatccat	cttgttcaat	catgcgaaac	gatcctcatc	ctgtctcttg	2520
atcagatctt	gatccccctg	gccatcagat	ccttgggcggc	aagaaagcca	tccagtttac	2580
tttgacaggg	ttcccaacct	taccagaggg	cgccccagct	ggcaattccg	gttcgtctgc	2640
tgcccaataa	accgcccagt	ctagctatcg	ccatgtaagc	ccactgcaag	ctacctgctt	2700
tctcttttgc	cttgccgtttt	cccttggtcca	gatagcccag	tagctgacat	tcattccggg	2760
tcagcacctg	ttctgcccag	tggtctttcta	cgtgtttccg	ttcttttagc	agcccttgcg	2820
ccctgagtcg	ttgcccgcagc	gtgaaagctt	tttgcaaaag	cctaggcctc	caaaaaagcc	2880
tcctcactac	ttctggaata	gctcagaggg	cgaggcggcc	taaataaaaa	aaattagtca	2940
gccatggggc	ggagaatggg	cggaactggg	cggaagttagg	ggcgggatgg	gcggagttag	3000
gggcccggac	atggttgctg	actaatggag	atgcatgctt	tgcatacttc	tgccgtgctg	3060
ggagcctggg	gactttccac	acctggttgc	tgactaattg	agatgcatgc	tttgcatact	3120
tctgcctgct	ggggagcctg	gggactttcc	acaccctaac	tgacacacat	tccacagccg	3180
gatctgcagg	acccaacgct	gcccagatg	cgccgcgtgc	ggctgctgga	gatggcggac	3240
gcgatggata	tgttcttgcca	agggttgggt	tgccgattca	cagttctccg	caagaattga	3300
ttggctccaa	ttcttgaggt	ggtgaatccg	ttagcgaggt	gccgccggct	tccattcagg	3360
tcgaggtggc	ccggctccat	gcaccgcgac	gcaacgcggg	gaggcagaca	aggtataggg	3420
cggcgcctac	aatccaatgc	aacctgttcc	atgtgtcgc	cgaggcgcat	aaatcgccgt	3480
gacgatcagc	ggtccaatga	tcgaagttag	gctggtaaga	gccgcgagcg	atccttgaag	3540
ctgtccctga	tggtcgatcat	ctacctgcct	ggacagcatg	gcctgcaacg	cggcatcccc	3600
atgccgcggg	aagcgagaag	aatcataatg	gggaaggcca	tcagccctcg	cgctcgccaac	3660
gccagcaaga	cgtagcccag	cgcgtcgggc	gcccattgcc	gcgataatgg	cctgcttctc	3720
gccgaaacgt	ttggtggcgg	gaccagtgc	gaaggcttga	gcgagggcgt	gcaagattcc	3780
gaataccgca	agcgacaggg	cgatcatcgt	cgcgctccag	cgaaagcggg	cctcgccgaa	3840
aatgaccgag	acgctgccc	gcacctgtcc	tacgagttgc	atgataaaga	agacagtcac	3900
aagtgcggcg	acgatagtca	tgccccgcgc	ccaccggaag	gagctgactg	ggttgaaggc	3960
tctcaagggc	atcggctcgac	gctctccctt	atgcgactcc	tgcattagga	agcagcccag	4020

tagtaggttg	aggccgttga	gcaccgccc	cgcaaggaat	ggtgcatgca	aggagatggc	4080
gccccacagt	cccccgcca	cgggcctgcc	accataccca	cgccgaaaca	agcgctcatg	4140
agccgaagt	ggcgagcccg	atcttcccca	tcggtgatgt	cgccgatata	ggcgccagca	4200
accgcacctg	tggcgccggt	gatgccggcc	acgatgcgtc	cgccgtagag	gatcttggca	4260
gtcacagcat	gcgcataatc	atgcttcgac	catgcgctca	caaagtaggt	gaatgcgcaa	4320
tgtagtacc	acatcgcat	cgctttccac	tgctctcgcg	aataaagatg	gaaaaatcaat	4380
ctcatggtaa	tagtccatga	aaatccttgt	attcataaat	cctccaggta	gctatatgca	4440
aattgaaaca	aaagagatgg	tgatctttct	aagagatgat	ggaatctccc	ttcagtatcc	4500
cgatgggtcaa	tgcgctggat	atgggataga	tgggaatatg	ctgattttta	tgggacagag	4560
ttgggaactg	ttcccaacta	aaatcatttt	gcacgatcag	cgactacga	actttaccca	4620
caaatagtca	ggtaatgaat	cctgatataa	agacagggtg	ataaatcagt	cttctacgcg	4680
catcgacgc	gcacaccgta	gaaagtcttt	cagttgtgag	cctgggcaaa	ccgttaactt	4740
tcggcggtt	tgctgtcgca	caggctcacg	tctaaaagga	aataaatcat	gggtcataaa	4800
attatcacgt	tgctcgcgcg	ggcgacggat	gttctgtatg	cgctgttttt	ccgtggcgcg	4860
ttgctgtctg	gtgatctgcc	ttctaaatct	ggcacagccg	aattgcgcga	gcttgggttt	4920
gctgaaacca	gacacacagc	aactgaatac	cagaagaana	atcactttac	ctttctgaca	4980
tcagaaggcg	agaaatttgc	cggtgaacac	ctgggtcaata	cgctgttttg	tgagcagcaa	5040
tattgcgctt	cgatgacgct	tggcggtgag	attgatacct	ctgctgcaca	aaaggcaatc	5100
gacgagctgg	accagcgcat	tcgtgacacc	gtctccttcg	aacttattcg	caatggagtg	5160
tcattcatca	aggacgcgcg	tatcgcaaat	ggtcgtatcc	acgcagcgcg	aatcgaaaca	5220
cctcagccgg	tgaccaatat	ctacaacatc	agccttggtg	tccagcgatg	tgagccagcg	5280
cagaacaagg	taaccgtcag	tgccgataag	ttcaaagtta	aacctgggtg	tgataccaac	5340
attgaaacgt	tgatcgaaaa	cgcgctgaaa	aacgctgctg	aatgtgcggc	gctggatgtc	5400
acaaagcaaa	tggcagcaga	caagaaagcg	attggtgaac	tggcttccta	tgtccgcacg	5460
gccatcatga	tggaaatgtt	ccccgggtgg	gttatctggc	agcagtgccg	tcgatagtat	5520
gcaattgata	attattatca	tttgcggtgc	ctttccggcg	atccgccttg	ttacggggcg	5580
gcgacctcgc	gggttttcgc	tatttatgaa	aattttccgg	tttaaggcgt	ttccgttctt	5640
cttcgctcata	acttaattgt	tttatttaaa	ataccctctg	aaaagaaagg	aaacgacagg	5700
tgctgaaagc	gagctttttg	gcctctgtcg	tttcccttct	ctgtttttgt	ccgtggaatg	5760
aacaatggaa	gtcaacaaaa	agcagctggc	tgacattttc	ggtgagagta	tccgtaccat	5820
tcagaactgg	cagggaacagg	gaatgcccgt	ctgcgagggc	ggtggcaagg	gtaatgaggt	5880
gctttatgac	tctgccgcgc	tcataaaatg	gtatgccgaa	agggatgctg	aaattgagaa	5940
cgaaaagctg	cgccgggagg	ttgaagaact	gcggcaggcc	agcgaggcag	atccacagga	6000
cgggtgtggt	gcctagtcga	cgttagtcga	tagtggctcc	aagttagcgaa	gcgagcagga	6060
ctgggcggcg	gcaaagcggt	cggacagtgc	tccgagaacg	ggtgcgcata	gaaattgcat	6120
caacgcata	agcgctagca	gcacgccata	gtgactggcg	atgctgtcgg	aatggacgat	6180
atcccgcaag	aggcccgcca	gtaccggcat	aaccaagcct	atgcctacag	catccagggt	6240
gacggtgccc	aggatgacga	tgagcgcat	gttagatttc	atacacgggtg	cctgactgcg	6300
ttagcaattt	aactgtgata	aactaccgca	ttaaagctta	tcgatgataa	gcggtcaaac	6360
atgagaattc	gcgccgcgtc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccatc	6420
cgtctgccgg	tgggtgtgtg	gcgtggctct	tgcggtctct	cgcccgacag	ctgcccccg	6480
cgcacttttc	tcagtggttc	gtgtctcgct	tgaccatggt	tgaggcgccc	ggttgtgccc	6540
tcacgtgttt	cactttggtc	tcccgtctgg	tgtgtgcacg	cccagagtcg	gtggatgtgg	6600
ccggtggcgt	tgcataccct	caggttgttc	tcctaggtgc	cgctgtttct	tgtaagcgct	6660
gaggtgctcc	tggagcggtc	tgccctccgt	gctccgtctg	ctgcttctga	gctgggtggg	6720
gcgctcccca	ttccctgggtg	gggggggtcga	ggagagaagg	gctgtgtgcc	ttcccgtttg	6780
tgtctgagaa	gcccgtgaga	cccgcgacta	gtacgcctgt	aggggcaaga	cccccttct	6840
tgctcgggtg	aggcgcccac	agtttctcga	gagactcatt	gcgtagggtc	ggtgctgagc	6900
ggtcgcggct	ggggttggaa	gggaccggtt	gcagggtctc	gctttcccgt	ggggagcttt	6960
gagaggcctg	gctttcgggg	ttcctgttgc	cgcagacccc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	tggttccctg	gcgatgcccc	cccgcgcggg	cgcccgctg	7080
ttggtcttct	gggttccctg	ccgcccctcag	gcctgcaccc	tctctcggtg	gcccgggctc	7140
gtcgggggtt	tgggtccgtc	acgggtctcga	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaaggggtg	cgggcttctt	cgcgcgcgca	ggggtctctc	ccgaatggtc	ccctggaggg	7260
ctcgccccct	gaccgcctcc	ggggagggat	gcgtttgtct	tctcgtctac	cgccggccgc	7320
ggcctccccg	ctccgagttc	gctcggcttg	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtggtgggt	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttcctcc	cctcctgagg	gccgccgtgc	ggacgggggtg	7500
tgggtaggcg	acggtgggct	cccggtccc	cacccgtctt	cccgtgcctc	accgctgcct	7560
tccgtcgctg	gcgtccctct	cgctcgcgtc	cacgactttg	gccgctcccg	cgacggcggc	7620
ctgcgcgcgc	cgtggtgcgt	gctgtgtgct	tctcgggctg	tgtggttgtg	tcgctcgcgc	7680
cccccttccc	cgcgcgcagc	ttcccacggc	tggcgaaatc	gcgggagtc	tccttccccct	7740
cctcgggggtc	gagagggtcc	gtgtctggcg	ttgattgata	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaacggctg	ttggccgcgt	ccggcgcgac	gtcggacgtg	gggaccact	7860
gcccgtcggg	ggtcttcgtc	ggtaggcac	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cgggtgtcgc	tcctcgggct	cccggggggc	cgtcgtgttt	cgggtcggtc	cggcgctgca	7980
ggtgtgggtg	gactgctcag	gggagtgggt	cagtggtgatt	cccgcgggtt	ttgcctcgcg	8040

tgccectgacc	gggtccgacgc	ccgagcggtc	tctcggtccc	ttgtgaggac	ccccctccgg	8100
gagggggccc	tttcggccgc	ccttgccgtc	gtcgccggcc	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgccgca	gccgggtctt	tttcctctct	ccccccctct	cctctgactg	8220
accogtggcc	gtgctgtcgg	accccccgca	tggggggcgg	cgggcacgta	cgcgctccgg	8280
cggtcaccgg	ggctctgggg	ggggggccgag	gggtaagaaa	gtcggctcgg	cgggcgggag	8340
gagctgtggt	ttggaggcgc	tcccgccccc	gcggccgtgg	cggtgtcttg	cgcggtcttg	8400
gagagggctg	cgtgcgaggg	gaaaaggttg	cccccgaggg	gcaaagggaa	agaggctagc	8460
agtggtcatt	gtcccgacgg	tgtgggtggtc	tgttggccga	gggtgcgtctg	ggggggtcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgcgt	gttggggcct	gccggagtgc	cgaggtgggt	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgggtg	tggcgggtgg	ggctccgggtc	8640
gatgtctacc	tccctctccc	cgaggtctca	ggccttctcc	gcgcgggtctc	tcggccctcc	8700
cctcgttcct	ccctctcgcg	gggttcaagt	cgctcgtcga	cctccccctc	tcogtccctc	8760
catctctcgc	gcaatggcgc	cgcccgagtt	caacgtgggt	tcgtcctccg	cctccgcttc	8820
tcgcccgggg	ctggccgctg	tccgggtctct	cctgcccgac	ccccgttggc	gtgggtctct	8880
ctcgccggct	tcgcccagctc	ctgggtctcgc	ccggagggtc	aggggggttc	ccgggtcccc	8940
gacgttgccg	ctcgctcgtg	tgtgcttggg	ggggggccgc	tgccggcctcc	gcccgcctgt	9000
gagcccccgc	cgcacccgcc	gggtgtcgggt	ttcggcgccg	gggtcagttgg	gccctggcgt	9060
tgtgtcgcgt	cgggagcgtg	tccgcctcgc	ggcggtcaga	cgcggtgtgc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgccccgc	ccgacccccg	cctgccccgc	ccgggtgggtg	9180
tcgttggtgt	ggggagtga	tgggtctacc	ggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcccgtgt	cgcgcttctc	tttcgcccaa	ccccccacgc	aaccaccac	cctgctctcc	9300
cggcccgggtg	cggctcgacgt	tccggctctc	ccgatgcgga	gggggttcggg	atthgtgccc	9360
gggacggagg	ggagagcggg	taagagaggt	ctcggagagc	tgtcccgggg	gcacgctcgg	9420
ggttggtttg	cgcgctcgtg	gtgctcgcgg	acgggttttg	tcggaccccg	acgggggtcgg	9480
tccggccgca	tgcactctcc	cgttccgcgc	gagcgcgcgc	ccgggtcacc	cccggtttgt	9540
cctcccgcca	ggctctccgc	cgcgcgcgc	tcctcctcct	ctctcgcgct	ctctgtcccc	9600
cctggctcgt	tcccacccc	gacgctccgc	tcggcgttcc	ttacctggtt	gatccgtcca	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatgggtc	attaaatcag	ttatgggtcc	tttgggtcgt	cgctcctctc	9780
tactctggat	aactgtggta	attctagagc	taatacatgc	cgacggggcg	tgacccccct	9840
tcccgggggg	ggatgctgct	atthtatcaga	tcaaaaacaa	cccggtgagc	tccctcccg	9900
ctccggccgg	gggtcgggcg	ccggcggctt	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccc	tggcggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgc	taccatgggtg	accacgggtg	acgggggaatc	aggggttcgat	tcgggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactctc	10140
cgacccgggg	aggtagtgc	gaaaaataac	aatacaggac	tctttcgagg	ccctgtaatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggtatccattg	gagggcaagt	ctgggtgcca	10260
cagcccggtt	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcggggcggtc	cgccgcgagg	cgagtcaccg	cccggtcccc	10380
ccccctgcct	ctcggcgccc	cctcgatgct	cctagctgag	tgtcccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	tcaggcccgga	gccccttgga	taccgagct	10500
aggaataatg	gaataggacc	gcgggttctat	tttgttggtt	ttcggaactg	aggccatgat	10560
taagagggac	ggccgggggg	attcgtattg	cgccgctaga	ggtgaaattc	ttggaccggc	10620
gcaagacgga	ccagagcgaa	agcattttgc	agaatgtttt	tcattaatca	agaacgaaag	10680
tcggagggtt	gaagacgac	agataccgtc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcggcg	gcgttattcc	catgacccgc	cgggcagctt	ccgggaaacc	aaagtctttg	10800
gggtccgggg	ggagtatggg	tgcaaaagctg	aaacttaaag	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcaccc	ggcccggaca	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctgggt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggtcg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgacgcgcgc	ctacactgac	tggctcagcg	tgtgcctacc	ctgcgcgggc	aggcgcggtg	11220
aaaccggtga	acccattcgt	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtccttgcc	ctttgtacac	11340
accgcccgtc	gctactaccg	attggatggg	ttagtgaggc	cctcggtatcg	gcccgcgcgg	11400
gggtcggccc	cggccctggc	ggagcgctga	gaagacgggtc	gaacttgact	atctagagga	11460
agtaaaaagtc	gtaacaagggt	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcgcgct	ggcccgctct	ccccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgcggtgcgt	ccgggtcccg	tcgcccgcgt	gtggagcgag	gtgtctggag	tgaggtgaga	11640
gaaggggttg	gtgggtcgg	tctgggtccg	gttgggaccg	cctccgattt	cccccccc	11700
tccctctccc	ctcgccggc	ctcgacctgc	ccacctacc	gcggcgcgcg	ctgctcgcg	11760
gcgtcttgcc	tctttcccg	ccggtctctc	cgtgtctacg	agggcggtga	cgctcgttac	11820
ggtttttgac	ccgtcccggg	ggcgttcgggt	cgtcggggcg	cgcgctttgc	tctcccggga	11880
cccatccccc	cccggtctct	ggcttttcta	cgttggtgg	ggcggttgct	gcgtgtgggg	11940
ggatgtgagt	gtcgcggtg	ggctcgcccg	tcccgatgcc	acgcttttct	ggcctcgcg	12000
gtcctccccc	ctcctgtccc	gggtacctag	ctgtcgcggt	ccggcgcgga	ggtttaagga	12060

ccccgggggg	gtcgccctgc	cgccccaggg	gtcggggggg	ggggggggcc	gtaggggaagt	12120
cggtcggtcg	ggcggtctct	cctcagactc	catgaccctc	ctccccccgc	tgccgcccgtt	12180
cccagaggcg	cggtcggtgt	gggggggtga	tgtctggagc	cccctcgggc	gccgtggggg	12240
cccagaccgc	ggcgccggct	tgcccgatct	ccgcgggtcg	gtcctgtcgg	tgccgggtcgt	12300
gggttcccg	gtcggtcccg	tggttttccg	ctcccgacc	tttttttttc	ctcccccca	12360
cacgtgtctc	gtttcggtcc	tgctggccgg	cctgaggcta	cccctcggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tcgttggggg	actgtgccc	cgtcagcacc	12480
cgtgagttcg	ctcacaccgc	aaataccgat	acgactctta	ggggtggatc	actcggtcgt	12540
tgctcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttccctc	cggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcacccgc	tgcggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	acccgggtcg	ggcctccgt	ctccgaagt	tcagacgtgt	12780
ggcggttgt	cggtgtggcg	cgcgcccg	cggtcgagg	cctggtctcc	cccgcgcatc	12840
cgcgctcgcg	gcttcttccc	gctccgccc	tcggcgccct	gcccgtgcac	cccgttccgt	12900
gcctcgcgct	ggcgccctcc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgcccc	12960
tggaaccca	ccgcgcccc	gtggcgccc	ggggtggcg	cgtccgcatc	tgctctgggt	13020
gaggttggcg	gttgaggggtg	tgctgcccgc	gaggtgggtg	tcgggtccct	gcggcccgcg	13080
ggttctcggg	gtggcggtcg	acgagggccg	gtcggtcgcc	tgcggtgggt	gtctgtgtgt	13140
gtttgggtct	tcgctggggg	gaggggggg	cgaccgctcg	cggggttggc	gcggctcgcc	13200
ggcgcccgcg	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatccccgg	tggtgttgcg	agggaggggt	tggtgtccc	cgtccgctcc	tcctccctc	13320
cctcggtggg	cgcttctcgg	ccgcacgcgg	ccgttagggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cttctctctc	cttctctctc	caccggcg	gtaccgctc	cgccgcccgc	13440
ccgcgggacg	ccgcggcgct	cgtgcgccga	tgcgagtcac	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gacccggtgc	gtcccggctt	ccctgggggg	gacccggcgt	13560
ctgtgggctg	tgcttcccgg	gggttgctgt	tgagtaagat	cctccacccc	cgccgcccct	13620
ccctcccgcc	ggcctctcgg	ggacccctcg	agacgggtcg	ccggctcgtc	ctccggtgce	13680
gccgggtgcc	gtctctttcc	cgcccgccct	ctcgctctct	tcttcccgcg	gctgggccc	13740
tgtccccctt	ttctgaccgc	gacctcagat	cagacgtggc	gacccgctga	atttaagcat	13800
attagtccag	ggagaaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcgaatc	cccgcgcgcg	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gacccactcc	ccggcgccgc	tcgtgggggg	cccaagtcc	tctgatcgag	gcccagcccc	13980
tggaagggtg	gagggccggt	gcggcccccg	cgcgccgggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaaactca	tctaaggcta	aataccggca	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagttaa	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaaaccgt	taagaggtaa	acgggtgggg	tcgcgcagat	ccgcccggag	14220
gattcaaccc	ggcgcccgcc	gtccggccgt	gcccgggtgt	cccggcggt	ctttcccgt	14280
ccccgttctt	cccgaccctt	ccacccgcgc	gtcgttcccc	tcttctctcc	cgcgctccgg	14340
gcctccggcg	gcgggcgcgg	gggttgggtg	gggtgtggcg	cgcgggccgg	gcccgggggt	14400
gggtcgcgcg	gggaccgccc	ccggccggcg	accggccgcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	tcggggacgg	tcggggacgg	ccgggaaagg	ccgggtggga	aggtgggtcg	14520
ggggggggcg	cgcgctctcag	ggcgcccgca	accacctcac	cccagatgtt	acagccctcc	14580
ggcgcgcgct	tcggcgcaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgctctcc	14640
ctctcccccc	gtccgcctcc	cgggcgggcg	tggggtgggg	ggcgggcccg	ccccccacc	14700
ggcgcgaccg	ctctcccacc	cccttccgtc	gcctctctcg	ggggccgggt	gggggggggg	14760
cggactgtcc	ccagtgcgcc	ccggcgctcg	tcggcgccgt	gggtcccggg	gggaccgtcg	14820
gtcacgcgct	tcccgcagaa	gcccagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgaccgctct	tgaaacacgg	accaaggagt	ctaaccgctg	cgcgagtcag	gggctcgtcc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggccccgccc	ggggggcccg	ggtgggatcc	15000
cgaggccctc	ccagtccgcc	gagggcgcac	caccggcccg	tctcgcccgc	cgcgccgggg	15060
aggtggagca	cgagcgtacg	cgltaggacc	cgaaaagtgg	tgaactatgc	ttgggcaggg	15120
cgaaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	ataggggcga	aagactaatc	gaacctcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gtagctgggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggctctgggg	ccgaaacgat	ctcaacctat	tctcaaaact	taaatgggta	15360
agaagcccgg	ctcgctggcg	tggaagccgg	cgtggaatgc	gagtgccctg	tgggccactt	15420
ttggttaagca	gaactggcgc	tgcgggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agaccccaga	aaaggtgttg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	ccggcccgct	gcccagtcg	gaacgggaacg	15660
ggacggggagc	ggcgccgaat	tcttgaagac	gaaagggcct	cgtgatacgc	ctatttttat	15720
aggttaagt	cagcataata	atggtttctt	agacgtcagg	tggtcacttt	cggggaaatg	15780
tgcgcggaac	ccctatttgt	ttatttttct	aaatacatct	aaatatgtat	ccgctcatga	15840
gacaataacc	ctgataaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	15900
atttccgtgt	cgcctttatt	cccttttttg	cggtattttg	cttccgtgtt	ttgctcacc	15960
agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	ggtgcacgag	tggtttacat	16020
cgaaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	cgccccgaag	aacgttttcc	16080

aatgatgagc	actttttaaag	ttctgctatg	tggcgcggtg	ttatcccggtg	ttgacgcccgg	16140
gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	gacttgggttg	agtactcacc	16200
agtcacagaa	aagcatctta	cggatggcat	gacagtaaga	gaattatgca	gtgctgccat	16260
aaccatgagt	gataacactg	cggccaactt	acttctgaca	acgatcggag	gaccgaaggga	16320
gctaaccgct	tttttgcaca	acatggggga	tcatgtaact	cgccttgatc	gttgggaacc	16380
ggagctgaat	gaagccatac	caaacgacga	gcgtgacacc	acgatgcctg	cagcaatggc	16440
aacaacgttg	cgcaaaactat	taactggcga	actacttact	ctagcttccc	ggcaacaatt	16500
aatagactgg	atggaggcgg	ataaagtgtg	aggaccactt	ctgcgctcgg	cccttccggc	16560
tggctgggtt	attgctgata	aatctggagc	cggtagcgtg	gggtctcgcg	gtatcattgc	16620
agcactgggg	ccagatggta	agccctcccg	tatcgtagtt	atctacacga	cggggagtca	16680
ggcaactatg	gatgaacgaa	atagacagat	cgctgagata	ggtgcctcac	tgattaaagca	16740
ttggtaactg	tcagaccaag	tttactcata	tatactttag	attgatttaa	aacttcattt	16800
ttaattttaa	aggatctagg	tgaagatcct	ttttgataat	ctcatgacca	aaatccctta	16860
acgtgagttt	tcgttccact	gagcgtcaga	ccccgtagaa	aagatcaaag	gatcttcttg	16920
agatcctttt	tttctgcgcg	taatctgctg	cttgcaaaaca	aaaaaaccac	cgctaccagc	16980
ggtgggttgt	ttgcccgatc	aagagctacc	aactcttttt	ccgaaggtaa	ctggcttcag	17040
cagagcgcag	ataccaaata	ctgtccttct	agtgtagccg	tagttaggcc	accacttcaa	17100
gaactctgta	gcaccgccta	catacctcgc	tctgctaata	ctgttaccag	tggctgctgc	17160
cagtggcgat	aagtcgtgtc	ttaccgggtt	ggactcaaga	cgatagttac	cggataaggc	17220
gcagcgtcgc	ggctggaacg	gggggttcgtg	cacacagccc	agcttggagc	gaacgacctt	17280
caccgaactg	agatacctac	agcgtgagct	atgagaaagc	gccacgcttc	cgaagggaga	17340
aaggcggaca	ggtatccggt	aagcggcagg	gtcggaaacg	gaga		17384

<210> 119

<211> 2814

<212> DNA

<213> Artificial Sequence

<220>

<223> pLITMUS38 Plasmid

<400> 119

gttaactacg	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccctta	tttgtttatt	60
tttctaaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctgggtg	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagtctt	360
gctatgtggc	gcggtattat	cccgtgttga	cgccggggca	gagcaactcg	gtcgcgcgat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggacat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccctgtag	aatggcaaca	acgttgcgca	aactattaac	720
tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	780
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggttttattg	ctgataaaatc	840
tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggttaagcc	900
ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	cttttagattg	atttaccocg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttgtt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aatcaaaaag	aatagcccga	gatagggttg	agtgttgttc	cagtttggaa	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggt	cgaggtgccc	taaagcacta	1380
aatcggaaacc	ctaaagggag	ccccgatttt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaaagc	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgctgcgcgt	aaccaccaca	cccgcgcgcg	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctagggtga	agatcccttt	tgataatctc	atgaccaaaa	tcccttaacg	tgagttttcg	1620
ttccactgag	cgtcagaccc	cgtagaaaag	atcaaaggat	cttctttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttgtttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaaataactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaactcct	ttaccagtgg	ctgctgccag	tggcgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	cgggtcgggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaaagcgcc	acgcttcccg	aaggggagaaa	ggcggacagg	2100

tatccggtaa	gcggcagggt	cgggaacagga	gagcgcacga	gggagcttcc	aggggggaaac	2160
gcctgggtatc	tttatagtc	tgctcgggtt	cgccacctct	gacttgagcg	tcgattttttg	2220
tgatgctcgt	caggggggcg	gagcctatgg	aaaaacgcc	gcaacgcggc	cttttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agtttagctca	ctcatttaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tggtgaattg	tgagcgggata	2400
acaatttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gatatctgga	2520
tccacgaatt	cgctagcttc	ggccgtgacg	cgtctccgga	tgtagaggca	tgcgctcgacc	2580
ctctagtcaa	ggccttaagt	gagtcgtatt	acggactggc	cgtcgtttta	caacgctcgtg	2640
actgggaaaa	ccctggcggt	acccaactta	atcgcccttg	agcacatccc	cctttcgcca	2700
gctggcgtaa	tagcgaagag	gcccgcaccg	atcgcccttc	ccaacagttg	cgagcctga	2760
atggcgcaatg	gcgcttcgct	tggttaataaa	gcccgccttcg	gcgggctttt	tttt	2814

<210> 120

<211> 2847

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attB Plasmid

<400> 120

gttaactacg	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgttttatt	60
tttctaata	cattcaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcgcca	ttttgccttc	ctgtttttgc	tcacccagaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccgtgttga	cgccggggcaa	gagcaactcg	gtcgcgcgat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tgcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccgttagc	aatggcaaca	acgttgcgca	aactattaac	720
tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	780
agttgcagga	ccactttctgc	gctcggccct	tcgggctggc	tggttttattg	ctgataaatc	840
tggaagcggg	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaaagc	900
ctcccgatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	cttttagattg	atttaccctg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttgtt	aatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aaatcaaaag	aatagcccga	gatagggttg	agtgtgtgtc	cagtttgga	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aaccatcacc	caaatacaag	tttttggggt	cgaggtgccg	taaagcacta	1380
aatcggaacc	ctaaagggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcggta	1500
cgctgcgcgt	aaccaccaca	cccgcgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctaggtga	agatcccttt	tgataatctc	atgacaaaa	tcctttaacg	tgagttttcg	1620
ttccactgag	cgtcagaccc	cgtagaaaag	atcaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttgtttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaaatactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaactctg	ttaccagtgg	ctgctgccag	tgccgataag	1920
tcgtgtctta	ccgggttggg	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	1980
tgaacggggg	gttcgtgcac	acagccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aaggagagaa	ggcggacagg	2100
tatccggtaa	gcggcagggt	cggaaacagga	gagcgcacga	gggagcttcc	aggggggaaac	2160
gcctgggtatc	tttatagtc	tgctcgggtt	cgccacctct	gacttgagcg	tcgattttttg	2220
tgatgctcgt	cagggggggc	gagcctatgg	aaaaacgcc	gcaacgcggc	cttttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agtttagctca	ctcatttaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tggtgaattg	tgagcgggata	2400
acaatttcac	acaggaaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gattgaagcc	2520
tgctttttta	tactaacttg	agcgaaatct	ggatccacga	attcgctagc	ttcggcggcg	2580
acgcgtctcc	ggatgtacag	gcatgcgtcg	accctctagt	caaggcctta	agtgagtcgt	2640
attacggact	ggccgtcggt	ttacaacgctc	gtgactggga	aaaccctggc	gttaccacaac	2700

ttaatcgccct	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	2760
ccgattacccc	ttcccaacag	ttgcgcagcc	tgaatggcga	atggcgcttc	gcttggtaat	2820
aaagcccgct	tggcggggct	ttttttt				2847

<210> 121
 <211> 4223
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pLIT38attBBSRpolyA2 Plasmid

<400> 121						
accatgaaaa	cattttaacat	ttctcaacaa	gatctagaat	tagtagaagt	agcgacagag	60
aagattacaa	tgcttttatga	ggataataaa	catcatgtgg	gagcggcaat	tcgtacgaaa	120
acaggagaaa	tcattttcggc	agtacatatt	gaagcgtata	taggacgagt	aactgtttgt	180
gcagaagcca	ttgcgattgg	tagtgcagtt	tcgaatggac	aaaaggattt	tgacacgatt	240
gtagctgtta	gacaccctta	ttctgacgaa	gtagatagaa	gtattcgagt	ggtaagtcct	300
tgtggatgtg	gtagggagtt	gatttcagac	tatgcaccag	attgttttgt	gttaatagaa	360
atgaatggca	agtttagtcaa	aactacgatt	gaagaactca	ttccactcaa	atataccgga	420
aattaaaagt	tttaccatac	caagcttggc	tgctgcctga	ggctggacga	cctcgcgagg	480
ttctaccggc	agtgcacaa	cgtcggcatc	caggaaacca	gcagcggcta	tccgcgcata	540
catgcccccg	aactgcagga	gtggggaggc	acgatggccg	ctttgggtccg	gatcctttgtg	600
aaggaaacctt	acttctgtgg	tgtgacataa	ttggacaaac	tacctacaga	gatttaaagc	660
tctaaggtaa	atataaaatt	tttaagtgt	taatgtgtta	aactactgat	tctaattgtt	720
tgtgtatttt	agattccaac	ctatggaact	gatgaatggg	agcagtgggtg	gaatgccttt	780
aatgaggaaa	acctgttttg	ctcagaagaa	atgccatcta	gtgatgatga	ggctactgct	840
gactctcaac	attctactcc	tccaaaaaag	aagagaaagg	tagaagaccc	caaggacttt	900
ccttcagaat	tgctaagttt	tttgagtcac	gctgtgttta	gtaatagaac	tcttgcttgc	960
tttgctattt	acaccacaaa	ggaaaaagct	gcactgctat	acaagaaaat	tatggaaaaa	1020
tattctgtaa	cttttataag	taggcataac	agttataatc	ataacatact	gttttttctt	1080
actccacaca	ggcatagagt	gtctgctatt	aataactatg	ctcaaaaaat	gtgtaccttt	1140
agctttttta	tttgtaaagg	ggtaataaag	gaatatttga	tgatatagtgc	cttgactaga	1200
gatcataatc	agccatacca	catttgtaga	ggttttactt	gctttaaaaa	acctcccaca	1260
cctccccctg	aacctgaaac	ataaaatgaa	tgcaatttgt	gttggttaact	tgttttattgc	1320
agcttataat	ggttacaaat	aaagcaatag	catcacaaat	ttcacaaaata	aagatccaga	1380
tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	cagagaagct	tggcgccagc	1440
cggcttcaat	tgacaggggc	ccactagtga	gtcgtattac	gtagcttggc	gtaatcatgg	1500
tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacaa	ttccacacaa	catacgagcc	1560
ggaagcataa	agtgtaaagc	ctggggtgcc	taatgagtga	gctaactcac	attacatgtg	1620
agcaaaaagg	cagcaaaaagg	ccaggaaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	1680
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtcaga	ggtggcgaaa	1740
cccgacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	1800
tggtccgacc	ctgcgcctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	1860
gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	taggtcgttc	gctccaagct	1920
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	1980
tcttgagtcc	aaaccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	2040
gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtggg	ggcctaacta	2100
cggctacact	agaagaacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	2160
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggttttt	2220
tgtttgcaag	cagcagatta	cgcgagaaaa	aaaaggatct	caagaagatc	ctttgatctt	2280
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	tggtcatgag	2340
attatcaaaa	aggatcttca	cctagatcct	tttacgcgcc	ctgtagcggc	gcattaagcg	2400
cggcgggtgt	gggtggttacg	cgcagcgtga	ccgctacact	tgccagcgcc	ctagcgccc	2460
ctcctttcgc	tttcttccct	tcctttctcg	ccacgttcgc	tttccccgto	aagctctaaa	2520
tcgggggctc	cctttagggt	tccgatttag	tgctttacgg	cacctcgacc	ccaaaaaact	2580
tgatttggtg	gatgggtcac	gtagtgggcc	atcgccctga	tagacgggtt	ttcgcccttt	2640
gacgttggag	tccacgttct	ttaatagtgg	actcttggtc	caaactggaa	caacactcaa	2700
ccctatctcg	ggctattctt	ttgatttata	agggattttg	ccgatttcgg	cctattgggt	2760
aaaaaatgag	ctgattttaac	aaaaatttaa	cgcgaaattt	aacaaaaatat	taacgtttac	2820
aatttaaata	tttgcttata	caatcttcct	gtttttgggg	cttttctgat	tatcaaccgg	2880
ggtaaatcaa	tctaaagtat	atatgagtaa	actttggtctg	acagttacca	atgcttaatc	2940
agtgaggcac	ctatctcagc	gatctgtcta	tttcgttcat	ccatagttgc	ctgactcccc	3000
gtcggttaga	taactacgat	acgggagggc	ttaccatctg	gccccagtcg	tgcaatgata	3060
ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccggaaag	3120
gccgagcgca	gaagtgggtc	tgcaacttta	tccgcctcca	tccagtcctat	taattgtgtg	3180
cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	gcaacggtgt	tgccattgct	3240

acaggcatcg	tgggtgtcacg	ctcgtcgttt	ggatggcctt	cattcagctc	cggttcccaa	3300
cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	aagcgggttag	ctccttcggt	3360
cctccgatcg	ttgtcagaag	taagttggcc	gcagtgttat	cactcatggt	tatggcagca	3420
ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	tgggtgagta	3480
tcaaccaagt	cattctgaga	atagtgtatg	cgccgaccga	gttgcctctt	cccggcgta	3540
acacgggata	ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	tggagaacgt	3600
tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	gatgtaacct	3660
actcgtgcac	ccaactgatc	ttcagcatct	tttactttca	ccagcgtttc	tgggtgagca	3720
aaaacaggaa	ggcaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	3780
ctcatactct	tcctttttca	atattattga	agcattttatc	aggggttattg	tctcatgagc	3840
ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	gggttccgcg	cacatttccc	3900
cgaaaagtgc	cacctgacgt	agttaacaaa	aaaaagcccg	ccgaagcggg	ctttattacc	3960
aagcgaagg	ccatttcgca	ttcaggctgt	ggaagggcga	gggaagggcga	tcgggtgcgg	4020
cctcttcgct	attacgccag	ctggcgaaag	ggggatgtgc	tgcaaggcga	ttaaagttgg	4080
taacgccagg	gttttcccag	tcacgacgtt	gtaaaacgac	ggccagtcgg	taatacgaact	4140
cacttaaggc	cttgactaga	gggtcgcgac	atgcctgtac	atccggagac	gcgtcacggc	4200
cgaagctagc	gaattcgtgg	atc				4223

<210> 122

<211> 2686

<212> DNA

<213> Artificial Sequence

<220>

<223> pUC18 Plasmid

<400> 122

tcgcgcgctt	cggatgatgac	ggtagaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccggggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtg	180
accatgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgcc	240
attcgcatt	caggctgcgc	aactgttggg	aaggcgatc	ggtagcggcc	tcttcgctat	300
tacgccagct	ggcgaaagg	ggatgtgctg	caaggcgatt	aagtgggta	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgcga	gcttgcatgc	ctgcaggctg	420
actctagagg	atccccgggt	accgagctcg	aattcgtaat	catggtcata	gctgtttcct	480
gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	540
aaagcctggg	gtgcctaatg	agtgccttaa	ctcacattaa	ttgcgttgcg	ctcactgcc	600
gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	gaatcgccca	acgcgcgggg	660
agaggcgggt	tgcgtattgg	gcgctcttcc	gcttcctcgc	tcactgactc	gctgcgctcg	720
gtcgttcggc	tgcggcgagc	ggtagcagct	cactcaaagg	cggttaatac	gttatccaca	780
gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaa	gccagcaaaa	ggccagggaac	840
cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggctcc	gccccctga	cgagcatcac	900
aaaaatcgac	gctcaagtca	gaggtggcga	aaccgcagag	gactataaag	ataccaggcg	960
tttcccctgc	cgagctccct	cgtgcgctct	cctgttcgca	ccctgcgct	taccgggata	1020
ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	atagctcacg	ctgtagggtat	1080
ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgacgaacc	cccgttcag	1140
cccagccgct	gcgccttacc	cggtaactat	cgtcttgagt	ccaaccgggt	aagacacgac	1200
ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	1260
gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	1320
atctgcgctc	tgttgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	1380
aaacaaacca	ccgctggtag	cgggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	1440
aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	ggtctgacgc	tcagtgggaa	1500
gaaaactcac	gttaagggat	tttggctcatg	agattatcaa	aaaggatcct	cacctagatc	1560
cttttaaat	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttggtct	1620
gacagttacc	aattgcttaat	cagtggagca	cctatctcag	cgatctgtct	atttcgttca	1680
tccatagttg	cctgactccc	cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	1740
ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	1800
ataaaccagc	cagccggaag	ggccgagcgc	agaagtgggtc	ctgcaacttt	atccgcctcc	1860
atccagttca	ttaattgttg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	1920
cgcaacgttg	ttgccattgc	tacaggcatc	gtggtgtcac	gctcgtcgtt	tggtatggct	1980
tcattcagct	cgggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	2040
aaagcgggta	cctccttcgg	tcctccgac	gttgctcagaa	gtaagttggc	cgcagtggtta	2100
tcactcatgg	ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	2160
ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	aatagtgat	gcggcgaccg	2220
agttgctctt	ggccggcgct	aataccgggc	aataccgggc	cacatagcag	aacttttaaa	2280
gtgctcatca	ttggaaaacg	ttcttcgggg	cgaaaactct	caaggatcct	accgctgttg	2340
agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	2400

accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataaagg	2460
gcgacacgga	aatggtgaat	actcatactc	ttcctttttc	aatattattg	aagcattttat	2520
cagggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaataa	2580
gggggtccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	tctaagaaac	cattattatc	2640
atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc		2686

<210> 123

<211> 8521

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXeGFPattB(6xHS4)2 Plasmid

<400> 123

tacggggcg	gggatccact	agttattaat	agtaaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgt	tcataatgcca	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccc	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtc	cgctattacc	atgggtcgag	gtgagcccca	cggtctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	gggggggggg	gggggcgcg	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcgggggc	aggcgagag	gtgcgcgcg	agccaatcag	agcgggcgcg	tccgaaagt	600
tccttttatg	gcgaggcggc	ggcgcgcgcg	gccctataaa	aagcgaagcg	cgcgcgggcg	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgcctcgc	gccgcctcgc	gccgcccgc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacgggc	cttctcctcc	780
gggctgta	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggtcggggg	ggtgctgctg	900
tgtgtgtgtg	cggtggggagc	gccgctgctg	gcccgcgctg	cccggcggtc	gtgagcgctg	960
cgggcgcggc	gcggggcctt	gtgctgctcg	cgtgtgctcg	aggggagcgc	ggccgggggc	1020
ggtgccccgc	agtgcggggg	ggctgcgagg	ggaacaaagg	ctgctgctcg	ggtgtgtgctg	1080
tgggggggtg	agcagggggg	gtggcgcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cgggcccgct	tcgggtgctg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	ccgtgcgggg	cggggggtgg	cgggcaggtg	gggtgcccgg	cgggcggggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggtc	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgctgag	agggcgctgag	1380
gacttccttt	gtcccaaatc	tggcgagacc	gaaatctggg	agggcgccgc	gcacccccctc	1440
tagcgggcgc	gggcgaagcg	gtgcgcgccc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgcgcccgcc	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cggggttcgg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctctacg	ctcctgggca	1680
acgtgctggt	tgttgtgtctg	tctcatcatt	ttggcaaaaga	attcgccacc	atggtgagca	1740
agggcgagga	gctgttcacc	gggggtggtg	ccatcctggt	cgagctggac	ggcgacgtaa	1800
acggccacaa	gttcagcggtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tgcccgtgcc	ctggccccacc	ctcgtgacca	1920
ccctgacct	cggcgtgcag	tgcttcagcc	gctaccccga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gccgaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcgcg	cctgagcaaa	gaccccaacg	agaagcgcca	tcacatgggtc	ctgctggagt	2400
tcgtgaccgc	cgccgggatac	actctcgcca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaaggtgg	tggtgggtgt	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccccttg	2580
agcatctgac	ttctggctaa	ttaaaggaaat	ttattttcat	tgcaatagtg	tgttggaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcattttaa	catcagaatg	2700
agtattttgg	ttagagtttg	gcaacatatg	ccatatgctg	gctgcatga	acaaaggtag	2760
ctataaagag	gtcatcagta	tatgaacacg	ccccctgctg	tcatttcctt	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atattttgtt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttctct	acatgtttta	ctagccagat	ttttcctcct	ctcctgacta	2940
ctccagtc	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gcccagctt	3000
ggcgtaatca	tggtcatagc	tgtttctctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaataag	tgagctaact	3120

cacattaatt	gcgttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcata	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgcgc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttatttat	3300
gcagaggccg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggctagt	ggatcccccg	ccccgtatcc	cccaggtgtc	tgcagggtca	aagagcagcg	3420
agaagcggtc	agaggaaagc	gatcccgctg	caccttcccc	gtgcccgggc	tgtccccgca	3480
cgctgccggc	tccgggatgc	ggggggagcg	cgggaccgga	gcggagcccc	gggaggctcg	3540
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	3600
gtccccgtga	gcggatccgc	ggccccgtat	cccccaggtg	tctgcaggct	caaagagcag	3660
cgagaagcgt	tcagaggaaa	gcgatcccg	gccaccttcc	ccgtgocccg	gctgtccccg	3720
cacgctgccc	gctcggggat	gcggggggag	gcgcggaccg	gagcggagcc	cggggcggct	3780
cgctgctgcc	ccctagcggg	ggagggacgt	aattacatcc	ctgggggctt	tggggggggg	3840
ctgtccccgt	gagcgggatcc	gcgggcccgt	atcccccagg	tgtctgcagg	ctcaaagagc	3900
agcgagaagc	gttcagagga	aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	3960
cgcacgctgc	cggctcgggg	atgcgggggg	agcgccggac	cggagcggag	ccccggggcg	4020
ctcgctgctg	ccccctagcg	ggggagggag	gtaattacat	ccctgggggc	tttggggggg	4080
ggctgtcccc	gtgagcggat	ccgcggcccc	gtatccccca	ggtgtctgca	ggctcaaaga	4140
gcagcgagaa	gcgttcagag	gaaagcgatc	ccgtgccacc	ttccccgtgc	ccgggctgtc	4200
cccgcacgct	cccggtccgg	ggatgcgggg	ggagcgccgg	accggagcgg	agccccgggc	4260
ggctcgctgc	tgcggcctag	cgggggaggg	acgtaattac	atccctgggg	gctttggggg	4320
ggggctgtcc	ccgtgagcgg	atcccgccgc	ccgtatcccc	cagggtgtctg	caggctcaaa	4380
gagcagcgag	aagcgttcag	aggaagcga	tcccgtgcc	ccttccccgt	gccccggctg	4440
tccccgcacg	aacttgagcg	ggggatgcgg	gggagcgcgc	ggaccggagc	ggagccccgg	4500
gcggctcgct	gctgccccct	agcgggggag	ggacgtaatt	acatccctgg	gggctttggg	4560
ggggggctgt	ccccgtgagc	ggatccgcgg	ccccgtatcc	cccaggtgtc	tgcagggtca	4620
aagagcagcg	agaagcgttc	agaggaagc	gatccccgtg	caccttcccc	gtgccggggc	4680
tgtccccgca	cgctgcccgc	tcggggatgc	ggggggagcg	ccggaccgga	gcggagcccc	4740
gggcggctcg	ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	4800
ggggggggct	gtccccgtga	gcggatccgc	ggggctgcag	gaattcgatt	gaagcctgct	4860
tttttatact	aacttgagcg	aaatcaagct	cctaggcttt	tgcacaaaagc	taacttgttt	4920
attgcagctt	ataatggtta	caaataaagc	aatagcatca	caaatttcac	aaataaagca	4980
ttttttcac	tgcattctag	ttgtggtttg	tccaaactca	tcaatgtatc	ttatcatgtc	5040
tggatccgct	gcattaatga	atcgcccaac	gcgcggggag	aggcgggtttg	cgtattgggc	5100
gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	cggcgagcgg	5160
tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	5220
agaacatgtg	agcaaaaagg	cagcaaaaag	ccaggaaccg	taaaaaggcc	gcgttgctgg	5280
cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccga	5340
ggtggcgaaa	cccgacagga	ctataaagat	accaggcgtt	tccccctgga	agctccctcg	5400
tgcgctctcc	tgtttccgac	ctgcgcgtta	ccggatacct	gtccgccttt	ctcccttcgg	5460
gggaagctag	ggtttctcaa	tgtcacgctc	gtaggtatct	cagttcgggtg	taggtcgttc	5520
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	5580
gtaactatcg	tottgagtc	aaccoggtaa	gacacgactt	atcgccactg	gcagcagcca	5640
ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	ttgaagtggg	5700
ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	5760
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	5820
gtggtttttt	tgtttgcaag	cagcagatta	cgcgacagaa	aaaaggatct	caagaagatc	5880
ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	5940
tggatcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aaatgaagtt	6000
ttaaatcaat	ctaaagtata	tatgagtaaa	cttggtctga	cagttaccaa	tgcttaatac	6060
gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagtggcc	tgactccccg	6120
tcgtgtagat	cgggagggct	cgggagggct	taccatctgg	ccccagtgtc	gcaatgatac	6180
cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	6240
ccgagcgag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattgttgcc	6300
gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	6360
caggcatcgt	gggtgtcacgc	tgcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	6420
gatcaaggcg	agttacatga	tcccccatgt	tgtgcacaaa	agcgggttagc	tccttcgggtc	6480
ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatgggt	atggcagcac	6540
tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	6600
caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	6660
tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	6720
cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	6780
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	6840
aaacaggaag	gcaaaaatgc	gcaaaaaagg	gaataagggg	gacacggaaa	tgttgaatac	6900
tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	6960
gatacatatt	tgaatgtatt	tagaaaaata	gcaaaaatagg	ggttccgcgc	acatttcccc	7020
gaaaagtggc	acctggtcga	cgggtatcgat	aagcttgata	tcgaattcct	gcagccccgc	7080
ggatccgctc	acggggacag	ccccccccc	aagccccag	ggatgtaatt	acgtccctcc	7140

```

ccccgctagggg ggcagcagcgg agccggcccggg gggtccgctc cgggtccggcg cccccccgc 7200
atccccgagc cggcagcgtg cgggggacagc ccggggcacgg ggaaggtggc acgggatcgc 7260
tttctcttga acgcttctcg ctgctctttg agcctgcaga cacttggggg atacggggcc 7320
gcggatccgc tcacggggac agcccccccc caaagccccc agggatgtaa ttacgtccct 7380
cccccgctag ggggagcagc cgagccggccc ggggtccgcg tccggtccgg cgctcccccc 7440
gcatccccga gccggcagcg tgcgggggaca gcccgggcac ggggaaggtg gcacgggatc 7500
gctttctctt gaacgcttct cgctgctctt tgagcctgca gacacctggg ggatacgggg 7560
ccgcggtatcc gctcacgggg acagcccccc cccaaagccc ccaggatgt aattacgtcc 7620
ctcccccgct agggggcagc agcagcgcgc ccgggggtcc gctccgggtc ggcgctcccc 7680
ccgcatcccc gagccggcag cgtgcgggga cagcccgggc acggggaagg tggcacggga 7740
tcgctttcct ctgaacgctt ctcgctgctc tttgagcctg cagacacctg ggggatacgg 7800
ggccgcggtat ccgctcacgg ggacagcccc ccccaaaagc cccagggat gtaattacgt 7860
ccctcccccg ctagggggca gcagcgagcc gcccggggtc ccgctccggt ccggcgctcc 7920
ccccgcaccc cggagccggc agcgtgcggg gacagcccgg gcacggggaa ggtggcacgg 7980
gatcgctttc ctctgaacgc ttctcgctgc tctttgagcc tgcagacacc tgggggat 8040
ggggcccgcg atccgctcac ggggacagcc ccgccccggg ctccgctccg gtccggcgct 8100
gtccctcccc cgttaggggg cagcagcgag ccgccccggg ctccgctccg aaggtggcac 8160
ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 8220
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 8280
acggggccgc ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 8340
acgtccctcc cccgctaggg ggcagcagcg agccggcccg gggtccgctc cgggtccggcg 8400
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc 8460
acgggatcgc tttctcttga acgcttctcg ctgctctttg agcctgcaga cacctggggg 8520
a 8521

```

<210> 124
 <211> 8851
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> p18EPOcDNA Plasmid

```

<400> 124
cagttgcccgg ccgggtcgcg cagggcgaaac tcccgcccc acgggtgctc gccgatctcg 60
gtcatggccgg gccggaggcg gtcccggaaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacacc caggccaggg tgttgtccgg caccacctgg 180
tcctggaccg cgtgatgaa caggggtcacg tcgtcccga ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagcccg tcgggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccttgatatac agatttcgct 360
caagttagta taaaaaagca ggcttcaatc cctgcagagaa gcttgatatac gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc cccccccaaa gcccccaggg atgtaattac 480
gtccctcccc cgttaggggg cagcagcgag ccgccccggg ctccgctccg gtccggcgct 540
ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 600
gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
acggggccgc ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 720
acgtccctcc cccgctaggg ggcagcagcg agccgcccgg ggctccgctc cgggtccggcg 780
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccgggcacgg ggaaggtggc 840
acgggatcgc tttctcttga acgcttctcg ctgctctttg agcctgcaga cacctggggg 900
atacggggcc gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960
ttacgtccct ccccgctag ggggacagc cgagccgccc ggggtccgc tccggtccgg 1020
cgctcccccc gcatccccga gccggcagcg tgcggggaca gcccgggcac ggggaaggtg 1080
gcacgggatc gctttctctt gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
ggatacgggg ccgcggtatcc gctcacgggg acagcccccc cccaaagccc ccaggatgt 1200
aattacgtcc ctccccgct agggggcagc agcgagccgc cgggggtccc gctccgggtc 1260
ggcgctcccc ccgcatcccc gagccggcag cgtgcgggga cagccccggc acggggaagg 1320
ggggatacgg gcccgcggt cgcctcacgg ggacagcccc tttgagcctg cagacacctg 1380
gttaattacgt ccctcccccg ctagggggca gcagcgagcc gcccggggt ccgctccgg 1500
ccggcgctcc ccccgcatcc ccgagccggc agcgtgcggg gacagcccg gacagccgg 1560
ggtggcacgg gatcgcttcc atccgctcac cgttaggggg cagcagcgag cgggacagcc 1620
tgggggatac gggggccggg cgttaggggg gcttctgctg ggggacagcc cccccccaaa gcccccagg 1680
atgtaattac gtccctcccc cgttaggggg cagcagcgag ccgccccggg ctccgctccg 1740
gtccggcgct cccccgcat ccccgagccg gcttctgct gctctttgag cctgcagaca 1800
aaggtggcac gggatcgctt ggatccacta gttattaata gtaatcaatt acggggtcat 1920
cctgggggat cccatatacg gagttccgcg ttacataact tacggtaaat ggcccgcccg 1980
tagttcatag

```

gctgaccgcc	caacgacccc	cgcccatgga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gacatttccat	tgacgtcaaat	gggtgggacta	tttacggtaa	actgcccact	2100
tggcagtaga	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaga	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagcccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atttatattat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcggggcg	ggcgggagag	tgcgggcgga	gccaatcaga	gcgggcgct	2460
ccgaaagt	ccttttatgg	cgaggcggg	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cgttgccctt	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccggcc	cggctctgac	tgaccgctt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgttggg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	ccttaaagggc	tccgggaggg	ccctttgtgc	ggggggggagc	ggctcggggg	2760
gtgctgtgct	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	cccgcgtgc	ccggcggtcg	2820
tgagcgctgc	ggggcgggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgcccgcgc	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgctgcggcg	2940
gtgtgtgctg	ggggggggga	gcaggggggtg	tggggcgggc	gggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggcct	cgggtgcggg	gctccgtgcg	3060
ggggctggcg	cggggctcgc	cgtgccgggc	gggggggtggc	ggcaggtggg	ggtgcccggc	3120
ggggcgggcg	cgccctcggg	cggggagggc	tcgggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcgggcg	tcgagggcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
ggggcgaggc	acttccctttg	tcccaaatct	ggcgggagccg	aaatctggga	ggcgccggcg	3300
caccctcctc	agcggggcg	ggcgaaagcgt	tgccggcgccg	gcaggaagga	aatggcgggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	cccccttctc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggctgcctt	cgggggggac	ggggcagggc	gggggttcggc	ttctggcggtg	3480
tgaccggcg	ctctagaatg	gggtgacag	aatgtcctgc	ctggctgtgg	cttctcctgt	3540
ccctgctgtc	gtccctcctg	ggcctcccag	tccctggcgc	cccaccacgc	ctcatctgtg	3600
acagccgagt	cctggagagg	tacctcttgg	aggccaagga	ggccgagaat	atcacgacgg	3660
gctgtgctga	acactgcagc	ttgaatgaga	atatcactgt	cccagacacc	aaagttaatt	3720
tctatgcctg	gaagaggatg	gaggtcgggc	agcaggccgt	agaagtctgg	cagggcctgg	3780
ccctgctgtc	ggaagctgtc	ctgcggggcc	aggccctgtt	ggtcaactct	tcccagccgt	3840
gggagccccc	gcagctgcat	gtggataaag	ccgtcagtg	ccttcgcagc	ctcaccactc	3900
tgcttcgggg	tctgggagcc	cagaaggaag	ccatctcccc	tccagatg	gcctcagctg	3960
ctccactccg	aacaatcact	gctgacactt	tcgcgaact	cttccgagtc	tactccaatt	4020
tcctccgggg	aaagctgaag	ctgtacacag	gggaggccctg	caggacaggg	gacagatgac	4080
gtacaagtaa	gaattcactc	ctcaggtgca	ggctgcctat	cagaaggtgg	tggctgggtg	4140
ggcactgccc	ctggctcaca	aataccactg	agatcttttt	ccctctgcca	aaaattatgg	4200
ggacatctag	aagcccccctg	agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	4260
tgcaatagtg	tgttggaatt	ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	4320
tcatttaaaa	catcagaatg	agtatttggg	ttagagtttg	gcaacatatg	ccatatgctg	4380
gctgccatga	acaaaggtgg	ctataaagag	tgatcatagta	tatgaaacag	ccccctgctg	4440
tccattcctt	attccataga	aaagccctga	cttgagggtta	gatttttttt	atattttgtt	4500
ttgtgttatt	tttttcttta	acatccctaa	aattttcctt	acatgtttta	ctagccagat	4560
ttttcctcct	ctctcagcta	ctcccagtc	tagctgtccc	tottctctta	tgaagatccc	4620
tcgacctgca	gccccagctt	gcctgcctgc	aggctgactc	tagtggatcc	cccgccccgt	4680
atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	aagcgatccc	4740
gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgtgc	cggctcgggg	atgcgggggg	4800
agcccgggac	cggagcgggg	ccccggggcg	ctcgtgctg	ccccctagcg	ggggaggggc	4860
gtaattacat	ccctgggggc	tttggggggg	ggctgtcccc	gtgagcggat	ccgcggcccc	4920
gtatccccc	ggtgtctgca	ggctcaaaga	gcagcgagaa	gcgttcagag	gaaagcgatc	4980
ccgtgccacc	ttccccgctg	cgggctgtc	cccgcacgct	gcccggctcgg	ggatgcgggg	5040
ggagcgccgg	accggagcgg	agccccgggc	ggctcgctgc	tgccccctag	cgggggaggg	5100
acgtaattac	atccctgggg	gctttggggg	ggggctgtcc	ccgtgagcgg	atccgcggcc	5160
ccgtatcccc	cagggtgtctg	caggctcaaa	gagcagcgag	aagcgttcag	aggaaagcga	5220
tccgctgcca	ccttccccgt	gcccgggctg	tccccgcacg	ctgcccggctc	ggggatgcgg	5280
ggggagcgcc	ggaccggagc	ggagccccgg	gcggctcgct	gctgccccct	agcgggggag	5340
ggacgtaatt	acatccctgg	gggctttggg	ggggggctgt	ccccgtgagc	ggatccgcgg	5400
ccccgtatcc	cccagggtgc	tgagggtcga	aagagcagcg	agaagcggtc	agaggaaagc	5460
gatcccgctg	caccttcccc	gtgcccgggc	tgtccccgca	cgctgcccggc	tcggggatgc	5520
ggggggagcg	ccggagccga	gcggagcccc	ggggggctcg	ctgctgcccc	ctagcggggg	5580
agggacgtaa	ttacatccct	ggggggcttg	ggggggggct	gtccccgtga	gcggatccgc	5640
ggccccgtat	ccccagggtg	tctgcaggct	caaagagcag	cgagaagcgt	tcagaggaaa	5700
gcgatcccg	gccaccttcc	ccgtgcccgg	gctgtccccg	cacgctgcgc	gctcggggat	5760
gcggggggag	cgccggagcc	gagcggagcc	ccggggcggt	cgctgctgcc	ccctagcggg	5820
ggagggacgt	aattacatcc	tcggggggct	tggggggggg	ctgtccccgt	gagcggatcc	5880
gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	5940
aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgctgc	cggctcgggg	6000

atgcggggggg	agcgccggac	cggagcggag	ccccgggggg	ctcgcctgctg	ccccctagcg	6060
gggggagggac	gtaattacat	ccctgggggg	tttggggggg	ggctgtcccc	gtgagcggat	6120
ccgcgggggct	gcaggaattc	gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	6180
ccgctcacaa	ttccacacaa	catacgagcc	ggaagcataa	agtgtaaaagc	ctgggggtgcc	6240
taatgagtga	gctaactcac	attaattgcg	ttgcgctcac	tgcccgcctt	ccagtcggga	6300
aacctgtcgt	gccagctgca	ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgctg	6360
attgggcgct	cttccgcctc	ctcgcctcact	gactcgctgc	gctcggctcgt	tcggctgcgg	6420
cgagcgggtat	cagctcactc	aaaggcggtg	atacgggtat	ccacagaatc	aggggataac	6480
gcaggaaaga	acatgtgagc	aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	6540
ttgctggcgt	ttttccatag	gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgctca	6600
agtcagagggt	ggcgaacccc	gacaggacta	taaagatacc	agggcgtttcc	ccctgggaagc	6660
tccctcgtgc	gctctcctgt	tccgaccctg	.ccgcttaccg	gatacctgtc	cgccctttctc	6720
ccttcgggaa	cgctggcgct	ttctcatagc	tcacgctgta	gggtatctcag	ttcgggtgtag	6780
gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	ttcagcccca	ccgctgcgcc	6840
ttatccggta	actatcgtct	tgagtccaac	ccggtgaagac	acgacttatc	gccactggga	6900
gcagccactg	cgaacaggat	tagcagagcg	aggatatgtag	gcggtgctac	agagttcttg	6960
aagtgggtggc	ctaactacgg	ctacactaga	aggacagtat	ttgggtatctg	cgctctgctg	7020
aagccagtta	ccttcggaaa	aagagttggg	agctcctgat	cgggcaaaaca	aaccaccgct	7080
ggtagcgggtg	gtttttttgt	ttgcaagcag	cagattacgc	gcagaaaaaaa	aggatctcaa	7140
gaagatcctt	tactcttttc	tacggggctc	gacgctcagt	ggaacgaaaa	ctcacgttaa	7200
gggatttttg	tcattgagatt	atcaaaaagg	atcttcacct	agatcctttt	aaattaaaaa	7260
tgaagtttta	aatcaatcta	aagtatatat	gagtaaaact	gggtctgacag	ttaccaatgc	7320
tttaacagtg	aatgcactat	ctcagcgatc	tgtctatttc	gttcattccat	agttgcctga	7380
ctccccgtcg	tgtagataac	tacgatacgg	gagggtttac	catctggccc	cagtgcctgca	7440
atgataccgc	gagaccacag	ctcaccggct	ccagattttat	cagcaataaa	ccagccagcc	7500
ggaagggccg	agcgcagaag	tggtcctgca	actttatccg	cctccatcca	gtctattaat	7560
tggtgcgggg	aagctagagt	aagtagttcg	ccagtttaata	gtttgcgcaa	cggtgttgcc	7620
attgctacag	gcacgtggtg	gtcacgctcg	tcgtttggta	tggtcttcatt	cagctccgggt	7680
tcccaacgat	caaggcgagt	tacatgatcc	cccatgttgt	gcaaaaaagc	ggtagctcc	7740
ttcggctctc	cgatcgttgt	cagaagtaag	ttggccgcag	tggtatcact	catgggttatg	7800
gcagcaactg	ataattctct	tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggt	7860
gagtactcaa	ccaagtcat	ctgagaatag	tgtagcgccg	gaccgagttg	ctcttgcccc	7920
gcgtcaatac	gggataatac	cgcgccacat	agcagaactt	taaaagtgtc	catcattgga	7980
aaacgttctt	cggggcgaaa	actctcaagg	atcttaccgc	tggtgagatc	cagttcgatg	8040
taacccactc	gtgcacccaa	ctgatcttca	gcattcttta	ctttcaccag	cgtttctggg	8100
tgagcaaaaa	caggaaggca	aaatgccgca	aaaaagggaa	taagggcgac	acggaaatgt	8160
tgaatactca	tactcttctc	ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	8220
atgagcggat	acataattga	atgtatttag	aaaaataaac	aaataggggt	tcgcgcgaca	8280
tttccccgaa	aagtgccacc	tgacgtagtt	aacaaaaaaa	agccccccga	agcgggcttt	8340
attaccaagc	gaagcggcat	tcgccattca	ggctgcgcaa	ctggtgggaa	gggcgatcgg	8400
tgccgggcctc	ttcgctatta	cgccagctgg	cgaaaagggg	atgtgctgca	agggcattaa	8460
gttgggtaac	gccagggttt	tcccagtcac	gacgttgtaa	aacgacggcc	agtcggtaat	8520
acgactcact	taaggccttg	actagagggt	cgacggtata	cagacatgat	aagatacatt	8580
gatgagtttg	gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	8640
tgtgatgcta	ttgctttatt	tgtaacattt	ataagctgca	ataaacaagt	tgggggtgggc	8700
gaagaactcc	agcatagagt	ccccgcgctg	gaggatcatc	cagccggcgt	cccggaatac	8760
gattccgaag	cccaaccttt	catagaaggc	ggcgggtggaa	tcgaaatctc	gtagcacgtg	8820
tcagtcctgc	tcctcgccca	cgaagtgcac	g			8851

<210> 125

<211> 10474

<212> DNA

<213> Artificial Sequence

<220>

<223> p18genEPO Plasmid

<400> 125

cagttgcccgg	ccgggtcgcg	cagggcgaaac	tccccccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgggaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccgcg	cacccacacc	caggccaggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggctcacg	tcgtcccggg	ccacaccggc	gaagtcgtcc	240
tcacaggaag	cccgggagaa	cccagagccgg	tcggtccaga	actcgaccgc	tcgggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacgggcactg	ctcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatc	gaattcctgc	420
agccccgcgg	atccgctcac	ggggacagcc	cccccccaaa	gccccagggg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	gtccggcgct	540

ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	aaggtggcac	600
gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgcctagg	ggcagcagcg	agccgcccgg	ggctccgctc	cggctccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggcacgg	ggaaggtggc	840
acgggatcgc	tttccctctga	acgcttctcg	ctgctctttg	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggac	agcccccccc	caaagccccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagccgccc	ggggctccgc	tccggctccgg	1020
cgctccccc	gcacccccga	gccggcagcg	tgccggggaca	gcccggggcac	ggggaagggtg	1080
gcacgggatc	gctttccctct	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcggtatcc	gctcacgggg	acagcccccc	cccaaaggcc	ccagggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcgagccgc	ccggggctcc	gctccggctcc	1260
ggcgctcccc	ccgcctcccc	gagccggcag	cgctcgggga	cagcccgggc	acggggaaagg	1320
tggcacggga	tcgctttcct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatagcg	ggcccgggat	ccgctcacgg	ggacagcccc	ccccaaaagc	ccccagggat	1440
gtaattacgt	ctcccccccg	ctagggggca	cgagcgagcc	gcccggggct	ccgctccggt	1500
ccggcgctcc	ccccgcctcc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
ggtggcacgg	gatcgctttc	ctctgaacgc	ttctcgctgc	tccttgagcc	tgacagacc	1620
tgggggatac	ggggcccgcg	atccgctcac	ggggacagcc	ccccccaaa	gccccagg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttatttaata	gtaatcaatt	acggggtcat	1920
tagttcctag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgctcg	1980
gctgaccgcc	caacgacccc	cgcccattga	cgtaataaat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gccagtaga	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcctc	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	ttcccccccc	ctcccccccc	ctccctttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcggggcga	ggcgggaggg	tgccggcgga	gccaatcaga	gcgccgcgct	2460
ccgaaagttt	ccttttatgg	cgaggcgggc	gcccggcgcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	cggtcgctg	cggtgccttc	gcccctgccc	ccgctccgcg	ccgcctcgcg	2580
ccgcccgcgc	cggtctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgtttggt	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	gttaaaaggc	tcggggaggg	ccttttgtgc	gggggggagc	ggctcggggg	2760
gtgctgtgct	gtgtgtgtgc	gtggggagcg	ccgcgtgctg	cccgcgctgc	ccggcggtcg	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgccgtccgc	gtgtgcgcga	ggggagcgcg	2880
gcccgggggg	gtgccccgcg	gtgccccggg	gctgagaggg	gaacaaaggc	tgccgtgcgg	2940
gtgtgtgcgt	ggggggggtg	tgaggggggt	tgggcgcgcc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggtct	cgggtgcggg	gctccgtgcg	3060
gggctgtggc	cggggctcgc	cgtgccgggc	gggggggtgg	ggcaggtggg	ggtgccccgg	3120
ggggcggggc	cgtctcgggc	cggggagggc	tggggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcggtcg	tcgagggcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcctttg	tcccaaactc	ggcgagcccg	aaatctggga	ggcgccgcgc	3300
cacccccctc	agcgggcgcg	ggcgaagcgg	tgccgcgcgc	gcaggaagga	aatgggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccttctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggtgcctt	cgggggggac	ggggcagggc	ggggttcggc	ttctggcgct	3480
tgaccggcgg	ctctagatgc	atgctcgagc	ggccgccaagt	gtgatggata	tctgcagaat	3540
tcgccccttc	tagaatgggg	gtgcacgggt	agtactcgcg	ggctgggcgc	ccccgccccg	3600
ccgggtccct	gtttgagcgg	ggatttagcg	ccccggctat	tggccaggag	gtggctgggt	3660
tcaaggaccg	gcgacttgct	aaggaccccc	gaagggggag	gggggtgggg	tgccctccacg	3720
tgccagcggg	gacttggggg	agtccttggg	gatggcaaaa	acctgacctg	tgaaggggac	3780
acagtttggg	ggttgagggg	aagaagggtt	gggggttctg	ctgtgccagt	ggagagggaag	3840
ctgataagct	gataacctgg	gcgctggagc	caccacttat	ctgccagagg	ggaagcctct	3900
gtcacaccag	gattgaagtt	tggccggaga	agtggatgct	ggtagctggg	ggtggggtgt	3960
gcacacggga	gcaggattga	atgaaggcca	gggaggcagc	acctgagtg	ttgcatggtt	4020
ggggacagga	aggacgagct	ggggcagaga	cgtggggatg	aagggaagct	tccttcaca	4080
gccacccttc	tcctcccccg	cctgactctc	agcctggcta	tcctgttctag	aatgtcctgc	4140
ctggctgtgg	cttctcctgt	ccctgctgtc	gctccctctg	ggcctccag	tcctggggcg	4200
cccaccacgc	ctcatctgtg	acagccgagt	cctggagagg	tacctcttgg	aggccaaagg	4260
ggccgagaat	atcacgggtg	gaccccttcc	ccagcacatt	ccacagaact	cacgctcagg	4320
gcttcaggga	actccctcca	gatccaggaa	cctggcactt	ggtttggggg	ggagtggga	4380
agctagacac	tgcccccccta	cataagaata	agctctggtg	ccccaaacca	tacctggaaa	4440
ctaggcaagg	agcaaagcca	gcagatccta	cggcctgtgg	gccagggccca	gagccttcag	4500
ggacccttga	ctccccgggc	tgtgtgcatt	tcagacgggc	tgtgctgaac	actgcagctt	4560

gaatgagaat	atcactgtcc	cagacaccaa	agttaatttc	tatgcctgga	agaggatgga	4620
ggtgagttcc	tttttttttt	tttttccttt	cttttggaga	atctcatttg	cgagcctgat	4680
tttggaatgaa	agggagaatg	atcgagggaa	aggtaaaatg	gagcagcaga	gatgaggctg	4740
cctggggcgca	gaggctcacg	tctataatcc	caggctgaga	tggccgagat	gggagaattg	4800
cttgagccct	ggagtttcag	accaacctag	gcagcatagt	gagatcccc	atctctacaa	4860
acatttaaaaa	aaattagtc	ggtgaagtgg	tgcattggtg	tagtcccaga	tatttggaag	4920
gctgagggcgg	gaggatcgct	tgagcccagg	aatttgaggc	tgcatgtgagc	tgtgatcaca	4980
ccactgcact	ccagcctcag	tgacagagtg	aggccctgtc	tcaaaaaaga	aaagaaaaaa	5040
gaaaaataat	gagggctgta	tggaatacat	tcattattca	ttcactcact	cactcactca	5100
ttcattcatt	cattcattca	acaagtctta	ttgcatacct	tctgtttgct	cagcttggtg	5160
cttggggctg	ctgaggggca	ggagggagag	ggtgacatgg	gtcagctgac	tcccagagtc	5220
cactccctgt	aggtcgggca	gcaggccgta	gaagtctggc	aggccctggc	cctgctgtcg	5280
gaagctgtcc	tgccggggcca	ggccctgttg	gtcaactcct	cccagccgtg	ggagcccctg	5340
cagctgcatg	tggaataaagc	cgtcagtgcc	cttcgcagcc	tcaccactct	gcttcgggct	5400
ctgggagccc	aggtagtag	gagcggacac	ttctgcttgc	cctttctgta	agaaggggag	5460
aagggctctg	ctaaggagta	caggaactgt	ccgtattcct	tccctttctg	tggcactgca	5520
gcgacctcct	gttttctcct	tggcagaagg	aagccatctc	ccctccagat	gcggcctcag	5580
ctgctccact	ccgaacaatc	actgctgaca	ctttccgcaa	actcttccga	gtctactcca	5640
atttccctcg	gggaaagctg	aagctgtaca	caggggaggc	ctgcaggaca	ggggacagat	5700
gacgtacaag	taagaattca	ctcctcaggt	gcaggctgcc	tatcagaagg	tgggtggctgg	5760
tgtggccaat	gccctggctc	acaaataacca	ctgagatcct	tttccctctg	ccaaaaatta	5820
tggggacata	atgaagcccc	ttgagcatct	gacttctggc	taataaagga	aatttatttt	5880
cattgcaata	gtgtgttggg	attttttgtg	tctctcactc	ggaaggacat	atgggagggc	5940
aaatcattta	aaacatcaga	atgagtattt	ggtttagagt	ttggcaacat	atgccatattg	6000
ctgggtgcca	tgaacaaagg	tggctataaa	gaggtcatca	gtatatgaaa	cagccccctg	6060
ctgtccattc	cttattccat	agaaaaagcct	tgacttgagg	ttagattttt	tttatatttt	6120
gttttgtgtt	atttttttct	ttaacatccc	taaaattttc	cttacatgtt	ttactagcca	6180
gatttttctc	cctctcctga	ctactcccag	tcatagctgt	ccctcttctc	ttatgaagat	6240
ccctcgacct	gcagcccaag	cttgcatgcc	tgcaggctcga	ctctagtggg	tcccccgccc	6300
cgtatccccc	aggctcaaa	aggctcaaa	agcagcgaga	agcgttcaga	ggaaagcgat	6360
cccgtgccac	cttcccctg	cccgggctgt	ccccgcacgc	tgccggctcg	gggatgcggg	6420
gggagcgccg	gaccggagcg	gagccccggg	cggtctcgctg	ctgcccccta	gccccgggag	6480
gacgtaatta	catcctctgg	ggcttttggg	gggggctgtc	cccgtgagcg	gatccgcggc	6540
cccgatatcc	ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	6600
atcccgatgc	accttccccg	tgccccgggt	gtccccgcac	gctgcccggct	cggggatgcg	6660
gggggagcgc	cgagccggag	cggagccccg	ggcggtctgc	tgctgcccc	tagcggggga	6720
gggacgtaat	tacatccctg	ggggctttgg	ggggggggctg	tccccgtgag	cggtccgcg	6780
gccccgtatc	ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaag	6840
cgatcccctg	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgccc	ctcggggatg	6900
cggggggagc	gcgggagccg	agcggagcgg	ccggcggtct	gctgtgccc	cctagcgggg	6960
gagggagcgt	attacatccc	tgggggcttt	gggggggggg	tgtccccctg	agcggatccg	7020
cggcccccgta	tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	7080
agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	7140
tgccgggggga	gcgcgggacc	ggagcggagc	cccggggggc	tcgctgctgc	cccctagcgg	7200
gggagggagc	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	7260
cgccggcccc	tatccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	7320
aaagcgatcc	cgtgccacct	tccccgtgoc	cgggctgtcc	ccgcacgctg	ccggctcggg	7380
gatgcggggg	gagcgccgga	ccggagcggg	gccccggggc	gctcgctgct	gccccctagc	7440
gggggagggg	cgtaattaca	tccctggggg	ctttgggggg	gggctgtccc	cgtgagcggg	7500
tccgcggccc	cgtatcccc	agggtgtctg	aggctcaaag	agcagcgaga	agcgttcaga	7560
ggaaagcgat	cccgtgccac	cttccccctg	cccgggctgt	ccccgcacgc	tgccggctcg	7620
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	7680
gcggggggag	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	7740
gatccgcggg	gctgcaggaa	ttcgtaatca	tggtcatagc	tgtttctctg	gtgaaattgt	7800
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctggggg	7860
gcctaataag	tgagctaact	cacattaatt	gcgtttgcgt	cactgcccgc	tttccagctg	7920
ggaaacctgt	cgtgccagct	gcattaatga	actcgccaac	gcgcggggag	aggcggtttg	7980
cgtattgctg	gctcttccgc	ttcctcgctc	atcgactcgc	tgcgctcggt	cgttcggctg	8040
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	8100
aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	8160
gcgttgctgg	gctttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	8220
tcaagtccga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	tccccctgga	8280
agctccctcg	tgcgctctcc	tggtccgacc	ctgcgcctta	ccggatacct	gtccgccttt	8340
ctccctctcg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcgggt	8400
taggtcgctc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	8460
gccttatccg	gtaactatcg	tcttgagctc	aacccggtaa	gacacgactt	atcgccactg	8520
gcagcagcca	ctggtaacag	gatttagcaga	gcgaggtatg	taggcgggtg	tacagagttc	8580

ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttgggtat	ctgcgctctg	8640
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	8700
gctggtagcg	gtgggttttt	tgtttgcaag	cagcagatta	cgcgagaaa	aaaaggatct	8760
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	8820
taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	8880
aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaaa	8940
tgcttaaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagtgtgc	9000
tgactccccg	tcgtgtagat	aactacgata	cgggaggggct	taccatctgg	ccccagtgtc	9060
gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	9120
gccggaaggg	ccgagcgag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	9180
aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttgtt	9240
gccatttgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	9300
gggtcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	agcggtttagc	9360
tccttcgggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatggtt	9420
atggcagcac	tgcatatatt	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	9480
gggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	9540
ccggcgctcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	9600
ggaaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	9660
atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	9720
gggttagcaaa	aaacaggaag	gcaaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	9780
tggtgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttatgtt	9840
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	gggtccgcgc	9900
acattttcccc	gaaaagtgcc	acctgacgta	gttaacaaaa	aaaagccgcg	cgaagcggcg	9960
tttattacca	agcgaagcgc	cattcgccat	tcaggctgcg	caactgttgg	gaaggcgcat	10020
cgggtgcgggc	ctcttcgcta	ttacgccagc	tgccgaaagg	gggatgtgct	gcaaggcgat	10080
taagttgggt	aacgccaggg	ttttcccagt	cacgacgttg	taaaacgacg	gccagtcctg	10140
aatacgactc	acttaaggcc	ttgactagag	ggtcgacggg	atcacagacat	gataagatac	10200
attgatgagt	ttggacaaac	cacaactaga	atgcagtgaa	aaaaatgctt	tatttgtgaa	10260
atttgtgatg	ctattgtctt	atttgtaacc	attataagct	gcaataaaca	agttgggggtg	10320
ggcgaaagaac	tcaggcatga	gatccccgcg	ctggaggatc	atccagcccg	cgtcccggaa	10380
aacgattccg	aagcccaacc	tttcatagaa	ggcgcggtg	gaatcgaaat	ctcgtagcac	10440
gtgtcagtc	tgctcctcgg	ccacgaagtg	cacg			10474

<210> 126

<211> 6119

<212> DNA

<213> Artificial Sequence

<220>

<223> pl8attBZeoeGFP Plasmid

<400> 126

cagttgccgg	ccgggtcgcg	cagggcggaac	tcccgccccc	acggetgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgggaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccg	cacccacacc	caggccagg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	caggggtcacg	tcgtcccgg	ccacaccggc	gaagtctgtc	240
tcacgaagt	cccgggagaa	cccgaagccgg	tcggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttctgt	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgggctg	caggtcgagg	420
gatcttcata	agagaagagg	gacagctatg	actgggagta	gtcaggagag	gaggaaaaat	480
ctggctagta	aaacatgtaa	ggaaaatttt	agggatgtta	aagaaaaaaa	taacacaaaa	540
caaaatataa	aaaaaatcta	acctcaagtc	aaggcttttc	tatggaataa	ggaatggaca	600
gcagggggct	gtttcatata	ctgatgacct	ctttatagcc	acctttgttc	atggcagcca	660
gcatatggca	tatgttgcca	aactctaacc	caaatactca	ttctgatgtt	ttaaatgatt	720
tgccctccca	tatgtccttc	cgagtggagag	acacaaaaaa	ttccaacaca	ctattgcaat	780
gaaaaataaat	ttcctttatt	agccagaagt	cagatgctca	aggggcttca	tgatgtcccc	840
ataattttttg	gcagaggggaa	aaagatctca	gtggatattg	tgagccagg	cattggccac	900
accagccacc	accttctgat	aggcagcctg	cacctgagga	gtgaattctt	acttgtagag	960
ctcgtccatg	ccgagagtga	tcccggcggc	ggtcacgaac	tccagcagga	ccatgtgac	1020
gcgcttctcg	ttggggctct	tgtccagggc	ggactgggtg	ctcaggtagt	ggttgtcggg	1080
cagcagcacg	gggcccgtcg	cgatgggggt	gttctgctgg	tagtggctcg	cgagctgcac	1140
gctgcccgtc	tcgatgttgt	ggcggtctct	gaagttcacc	ttgatgccgt	tcctctgctt	1200
gtcggccatg	atatagacgt	tgtggctgtt	gtagttgtac	tccagcttgt	gccccaggat	1260
gttgccgttc	tccttgaagt	cgatgccctt	cagctcgatg	cggttcacca	gggtgtcgcc	1320
ctcgaaacttc	acctcttgta	gggtccttga	gttgcgctcg	tccttgaaga	agatgggtgc	1380
ctcctggacg	tagccttcgg	gcatggcgga	cttgaagaag	tcgtgctgct	tcatgtgggtc	1440
ggggtagcgg	ctgaagcact	gcacgccgta	ggtcagggtg	gtcacgagg	tgggcccagg	1500

cacggggcagc	ttgccgggtgg	tgcagatgaa	cttcagggtc	agcttgccgt	aggtgggcatc	1560
gcectcgccc	tcgcccggaca	cgctgaactt	gtggccggtt	acgtcgccgt	ccagctcgac	1620
caggatgggc	accaccccg	tgaacagctc	ctcgcccttg	ctcaccatgg	tggcgaattc	1680
tttgccaaaa	tgatgagaca	gcacaacaac	cagcacgttg	cccaggagct	gtaggaaaaa	1740
gaagaaggca	tgaacatggt	tagcagaggc	tctagagccg	ccggtcacac	gccagaagcc	1800
gaaccccgcc	ctgcccgcgt	ccccccgaag	gcagccgtcc	ccctgcggca	gccccgaggc	1860
tggagatgga	gaaggggacg	gcggcgcgcc	gacgcacgaa	ggccctcccc	gcccatttcc	1920
ttcctgccc	cgccgcaccg	cttcgcccgc	gcccgcctaga	gggggtgccc	cggcgcctcc	1980
cagatttcgg	ctccgccaga	tttgggacaa	aggaagtccc	tgcgccctct	cgcacgatta	2040
ccataaaagg	caatggctgc	ggctcgccgc	gcctcgacag	ccgcccggcg	tccggggccg	2100
ccgcgcccc	cccccgagcc	ctccccggcc	cgaggcgggc	ccgccccgcc	cggcaccccc	2160
acctgcgcgc	acccccgcgc	cggcacggcg	agccccgcgc	cacgccccgc	acggagcccc	2220
gcacccgaag	ccgggcccgtg	ctcagcaact	cggggagggg	ggtgcagggg	gggggttacag	2280
cccgaccgcc	gcgcccacac	cccctgctca	cccccccacg	cacacacccc	gcacgcagcc	2340
tttgttcccc	tgcgagcccc	cccgaccgcg	ggggcaccgc	ccccggccgc	gctccccctc	2400
cgcacacgcg	gagcgacaaa	agccccgcgc	cgcgcccgca	gcgctcacag	ccgcggggca	2460
gcgcgggccg	cacgcggcgc	tccccacgca	cacacacacg	cacgcacccc	ccgagccgct	2520
cccccccgca	caaaaggccc	tccccgagcc	ctttaaggct	ttcacgcagc	cacagaaaag	2580
aaacgagccg	tcattaaacc	aagcgcta	tacagcccgc	aggagaaggg	ccgtcccccc	2640
cgctcacctt	tggtgagta	gcggtcagtc	agagccgggg	cgggcgccgc	gaggcgccgc	2700
ggagcggggc	acggggcgaa	ggcaacgcag	cgactcccgc	ccgcccgcgc	cttcgccttt	2760
tataggggcg	ccgcgcggcg	cgccctcgcca	taaaaggaaa	ctttcgagc	gcgcccgtct	2820
gattggctgc	cgccgcacct	ctccgcctcg	ccccgcccc	ccctcgccc	cgccccgcc	2880
cgccggggcg	gcgccccccc	cccccccgcc	cccctcgctg	cacaaaataa	ttaaaaataa	2940
aataaataca	aaattggggg	tggggagggg	ggggagatgg	ggagagtga	gcagaacgtg	3000
gggtcacctt	cgacccatgg	taatagcgat	gactaatacg	tagatgtact	gccaagttagg	3060
aaagtcccat	aaggtcatgt	actgggcata	atgccaggcg	ggccatttac	cgctattgac	3120
gtcaataggg	ggcgtacttg	gcatatgata	cacttgatgt	actgccaagt	gggcagttta	3180
ccgtaaatag	tccaccattt	gacgtcaatg	gaaagtccct	attggcgtaa	ctatgggaac	3240
atacgtcatt	attggagcca	atgggcgggg	gtcggtgggc	ggtcagccag	gcggggcatt	3300
taccgtaagt	tatgtaacgc	ggaactccat	atatgggcta	tgaactaatg	accccgta	3360
tgattactat	taataactag	aggatccccg	ggtaccgagc	tcgaattcgt	aatcatggct	3420
atagctgttt	cctgtgtgaa	attgttatcc	gctcacaaat	ccacacaaca	tacgagcccg	3480
aagcataaag	tgtaaacgct	ggggtgccta	atgagtgagc	taactcacat	taattgcgtt	3540
gcgctcactg	cccgctttcc	agtcgggaaa	cctgtcgctg	cagctgcatt	aatgaatcgg	3600
ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctct	tccgcttcct	cgctcactga	3660
ctcgctcgcg	gggtgcgttc	ggctgcggcg	agcgggtatca	gctcactcaa	aggcggtaat	3720
acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3780
aaaggccagg	aaccgtaaaa	aggccgcggt	gctggcggtt	ttccataggc	tccgcccccc	3840
tgacgagcat	cacaaaaatc	gacgtcaag	ttagaggtgg	cgaaacccga	caggactata	3900
aagataccag	gcgtttcccc	ctggaagctc	cctcgtcgcg	tctcctgttc	cgacctgcc	3960
gcttaccgga	tacctgtccg	cctttctccc	ttcggaagc	gtggcgcttt	ctcatagctc	4020
acgctgtagg	tatctcagtt	cggtgtaggt	cgctcgctcc	aagctgggct	gtgtgcacga	4080
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtcacaacc	4140
ggtaagacac	gacttatcgc	cactggcagc	agccactggt	aacaggatta	gcagagcgag	4200
gtatgtaggc	ggtgctacag	agttcttgaa	gtggtggcct	aactacggct	acactagaag	4260
gacagtattt	ggtatctgcg	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4320
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcggtggt	ttttttgttt	gcaagcagca	4380
gattacgcgc	agaaaaaaag	gatctcaaga	agatcccttg	atcttttcta	cggggtctga	4440
cgctcagtg	aacgaaaaact	caggttaagg	gatttttggt	atgagattat	caaaaaggat	4500
cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	tcaatctaaa	gtatatatga	4560
gtaaacttgg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4620
tctatttcgt	tcatccatag	ttgcctgact	ccccgtcggt	tagataacta	cgatacggga	4680
gggctttacca	cttgccccca	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4740
agattttatca	gcaataaaacc	agccagccgg	aagggccgag	cgcagaagtg	gtcctgcaac	4800
tttatccgcc	tccatccagt	ctattaattg	ttgccgggaa	gctagagtaa	gtagttcgcc	4860
agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	atcgtggtgt	cacgctcgct	4920
gtttgggtatg	gcttcattca	gctccggttc	ccaacgatca	aggcgagtta	catgatcccc	4980
catgttgctgc	aaaaaagcgg	ttagctcctt	cggtcctccg	atcgttgtca	gaagtaagtt	5040
ggcgcgagtg	ttatcactca	tggttatggc	agcactgc	aattctctta	ctgtcatg	5100
atccgtaaga	tgcttttctg	tgactggtga	gtactcaacc	aagtcattct	gagaatagtg	5160
tatgcggcga	ccgagttgtg	cttgccggcg	gtcaatacgg	gataataccg	cgccacatag	5220
cagaacttta	aaagtgtctca	tcatgggaaa	acgttcttcg	gggcgaaaa	tctcaaggat	5280
cttaccgctg	ttgagatcca	gttcgatgta	acccactcgt	gcacccaact	gatcttcagc	5340
atcttttact	tttctgggtg	tttctgggtg	agcaaaaaaca	ggaaggcaaa	atgccgcaaa	5400
aaagggaata	agggcgacac	ggaaatgttg	aataactcata	ctcttccttt	ttcaatatga	5460
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatac	atatttgaat	gtatttagaa	5520

aaataaaca	ataggggttc	cgcgcacatt	tccccgaaaa	gtgccacctg	acgtagtta	5580
caaaaaaaag	cccgccgaag	cgggctttat	taccaagcga	agcgccattc	gccattcagg	5640
ctgcgcaact	gttgggaagg	gcgatcgggtg	cgggcctctt	cgctattacg	ccagctggcg	5700
aaagggggat	gtgctgcaag	gcgattaagt	tgggtaacgc	cagggttttc	ccagtcacga	5760
cggtgtaaaa	cgacggccag	tccgtaatac	gactcactta	aggccttgac	tagagggctg	5820
acggtatata	gacatgataa	gatacattga	tgagtttgga	caaaccacaa	ctagaatgca	5880
gtgaaaaaaa	tgctttat	gtgaaatttg	tgatgctatt	gctttatttg	taaccattat	5940
aagctgcaat	aaacaagttg	gggtgggcga	agaactccag	catgagatcc	ccgcgctgga	6000
ggatcatcca	gcccgcgtcc	cggaaaacga	ttccgaagcc	caacctttca	tagaaggcgg	6060
cgggtggaatc	gaaatctcgt	agcacgtgtc	agtcctgctc	ctcggccacg	aagtgcacg	6119

<210> 127

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXLamInt Plasmid (Wildtype Integrase)

<400> 127

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacgggtaaa	tggccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgt	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccc	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtcac	cgctattacc	atgggtcgag	gtgagcccca	cgctctgctt	cactctcccc	420
atctccccc	cctccccacc	cccaattttg	tattttat	ttttttaatt	attttgtgca	480
gcatggggg	cggggggggg	gggggcgcgc	gccaggcggg	gcggggcggg	gagggggcg	540
gggcggggcg	aggcggagag	gtgcggcggc	agccaatcag	agcggcgcg	tccgaaagtt	600
tccttttatg	gcgagggcgg	ggcggcgcg	gccctataaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgccct	cgcgccgtgc	cccgctccgc	gccgcctcgc	gccgcccgc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtagcggg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgctgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggctcgggg	gggtgcgtgc	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgtcg	cccggcggtc	gtgagcgctg	960
cgggcgcggc	gtggggcctt	gtgcgctccg	cggtgtgcgc	aggggagcgc	ggccgggggc	1020
gggtgcccc	gggtgcgggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgc	gggtgtgtgc	1080
tggggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggtc	tcgggtgcgc	ggctccgtgc	ggggcggtgc	1200
gcggggctcg	ccgtgcgggg	cggggggtgg	cggcaggtgg	gggtgcgggg	cggggcgggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gcccgggagc	gcccgggggt	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgaggg	1380
gacttccttt	gtccccaaatc	tggcggagcc	gaaatctggg	aggcgccgc	gcacccccctc	1440
tagcggggcg	gggcgaagcg	gtgcggcgcc	ggcagggaag	aaatggcgcg	ggagggcctt	1500
cgtgcgtgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cgggggttcg	cgggggttcg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttcatt	ccttcttctt	tttctacag	ctcctgggca	1680
acgtgctggt	tggtgtgctg	totcatcatt	ttggcaaaag	attcatggga	agaaggcgaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaggacgggt	aaagagtttg	gattaggcag	agacaggcga	atcgcaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaag	cctctgacag	1920
cgagaatcaa	cagtataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tcctggccag	cagaggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgcgg	2100
caatgctcaa	tggatacata	gacgagggca	aggcggcgct	agccaagtta	atcagatcaa	2160
cactgagcga	tgcatcccca	gaggcaatag	ctgaaggcca	tataacaaca	aaccatgtcg	2220
ctgccactcg	cgcagcaaaa	tcagaggtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaaaaattta	tcaagcagca	gaatcatcac	catgttggct	cagacttgca	atggaaactg	2340
ctgttggttac	cgggcaacga	gttggtgatt	tatgcgaaat	gaagtgggtc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattgccatc	ccaacagcat	2460
tgcatattga	tgctctcgga	atatcaatga	aggaaacact	tgataaatgc	aaagagattc	2520
ttggcgggaga	aaccataatt	gcattacttc	gtcgcgaacc	gctttcatcc	ggcacagtat	2580
caagggtattt	tatgcgcgca	cgaaaagcat	caggctcttc	cttcgaaggg	gatccgccta	2640
cctttcacga	ttgtcgcagt	gtgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgctcaaca	tcttctcggg	cataagtcgg	acaccatggc	atcacagtat	cgtgatgaca	2760
gaggcagggga	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820

```

cctatcagaa ggtggtggct ggtgtggcca atgccctggc tcacaaatac cactgagatc 2880
tttttccctc tgccaaaaat tatgggggaca tcatgaagcc ccttgagcat ctgacttctg 2940
gctaataaag gaaatattatt ttcatgtgcaa tagtgtgttg gaatTTTTTg tgtctctcac 3000
tcggaaggac atatgggagg gcaaatcatt taaaacatca gaatgagtat ttggtttaga 3060
gtttggcaac atatgccata tgctgggtgc catgaacaaa ggtgggtata aagaggatcat 3120
cagtatatga aacagccccc tgctgtccat tccttattcc atagaaaagc cttgacttga 3180
ggtttagattt tttttatatt ttgttttgtg ttattttttt ctttaacatc cctaaaattt 3240
tccttacatg ttttactagc cagatttttc ctccctcctc gactactccc agtcatagct 3300
gtccctcttc tcttatgaag atccctcgac ctgcagccca agcttggcgt aatcatggtc 3360
atagctgttt cctgtgtgaa attgttatoc gctcacaatt ccacacaaca tacgagccgg 3420
aagcataaag tgtaaaagcct ggggtgccta atgagtgagc taactcacat taattgcgtt 3480
gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagcggatcc gcactcaat 3540
tagtcagcaa ccatagtccc gccctaact ccgcccaccc cgcccctaac tccgcccagt 3600
tccgcccatt ctccgccccca tggctgacta atttttttta tttatgcaga ggccgaggcc 3660
gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctaggcttt 3720
tgcaaaaagc taacttgttt attgcagctt ataattggtt caaataaaagc aatagcatca 3780
caaattttcac aaataaagca tttttttcac tgcattctag ttgtggtttg tccaaactca 3840
tcaatgtatc ttatcatgtc tggatccgct gcattaatga atcggccaac gcgcggggag 3900
aggcggtttg cgtattgggc gctcttcgct ttccctcgctc actgactcgc tgcgtcgggt 3960
cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg taataacggg 4020
atcaggggat aacgcaggaa agaacatgtg agcaaaaagg cagcaaaaagg ccaggaaccg 4080
taaaaaggcc gcgttgctgg cgtttttcca taggtccgc cccctgacg agcatcaca 4140
aaatcgacgc tcaagtcaaa ggtggcgaaa cccgacagga ctataaagat accaggcggt 4200
tccccctgga agctccctcg tgcgtctccc tggtccgacc ctgcccgtta ccggatacct 4260
gtccgccttt ctcccttcgg gaagcgtggc gctttctcaa tgcctcacgt gtaggatatc 4320
cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc 4380
cgaccgctgc gccttatccg gtaactatcg tcttgagtc aaaccggtaa gacacgactt 4440
atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcgggtg 4500
tacagatctc ttgaagtggg ggctaacta cggctacact agaaggacag tatttgggt 4560
ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa 4620
acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgagaaa 4680
aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaaaca 4740
aaactcacgt taagggatth ttgtcatgag attatcaaaa aggatcttca cctagatcct 4800
tttaaatata aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa cttggtctga 4860
cagttaccaa tgcttaata gtgaggcacc tatctcagcg atctgtctat ttcgttcatc 4920
catagtggc tgactcccc cgtgtagat aactacgata cgggaggggt taccatctgg 4980
ccccagtgc gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat 5040
aaaccagcca gccggaaggg ccgagcgcag aagtggctct gcaactttat ccgcctccat 5100
ccagtctatt aattggtgccc ggaagctag agtaagtagt tcgccagttat atagtttgcc 5160
caacgttggt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc 5220
attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa 5280
agcgggttagc tccttcggtc ctccgatcgt tgctagaagt aagttggccg cagtgttato 5340
actcatgggt atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt 5400
ttctgtgact ggtgagtagt caaccaagtc attctgagaa tagtgtatgc ggcgaccgag 5460
ttgctcttgc ccggcgtaaa tacgggataa taccgcgcca catagcagaa ctttaaaagt 5520
gctcatcatt ggaaaacggt cttcggggcg aaaactctca aggatcttac cgctgttgag 5580
atccagttcg atgtaaccca ctcggtgcacc caactgatct tcagcatctt ttactttcac 5640
cagcgtttct ggggtgagcaa aaacaggaag gcaaaaatgcc gaataagggc 5700
gacacggaaa tggtgaatac tcatactctt cctttttcaa tattattgaa gcatttatca 5760
gggttatgt ctcatgagc gatacatatt tgaattgatt tagaaaaata aacaaatagg 5820
ggttccgcgc acatttcccc gaaaagtgcc acctg 5855

```

<210> 128

<211> 303

<212> DNA

<213> Artificial Sequence

<220>

<223> Human FER-1 Promoter

<400> 128

```

tccatgacaa agcacttttt gagcccaagc ccagcctagc tcgagctaaa cgggcacaga 60
gacgccaccg ctgtcccaga ggcagtcggc taccggtccc cgctcccag ctccgccaga 120
gcgcgcgagg gctccagcg gccgccctc cccacagca ggggcgggt cccgcgcca 180
ccggaaggag cgggctcggg gcggggcgcg ctgattggcc ggggcgggct tgacgcggac 240
gcggctataa gagaccacaa gcgacccgca gggccagacg ttcttcgccc agagtcgggt 300
acc 303

```

<210> 129
 <211> 6521
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pIRES-BSR Plasmid

```

<400> 129
tcaatatttg ccattagcca tattattcat tgggttatata gcataaatca atattggcta 60
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120
aatatgaccg ccatgtttggc attgattatt gactagttat taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
gcctggctga ccgccccaacg acccccgcgc attgacgtca ataatgacgt atgttcccat 300
agtaacgcgc ataggggactt tccattgacg tcaatgggtg gagtatttac ggtaaaactgc 360
ccacttggca gtacatcaag tgtatcatat gccaaagtccg ccccttattg acgtcaatga 420
cggtaaatgg cccgcctggc attatgccc gtacatgacc ttacggggact ttcctacttg 480
gcagtacatc tacgtattag tcatcgctat taccatgggt atgcggtttt ggcagtacac 540
caatggggcgt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600
caatggggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactg 660
cgatcgcccg ccccggttgac gcaaatgggc ggtaggcgtg tacggtggga ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggg agtttatcac 780
agttaaattg ctaacgcagt cagtgttctt gacacaacag tctcgaactt aagctgcagt 840
gactctctta aggtagcctt gcagaagttg gtcgtgaggc actgggcagg taagtatcaa 900
ggttacaaga caggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgtttc tgataggcac ctattggtct tactgacatc cactttgcct ttctctccac 1020
aggtgtccac tcccagttca attacagctc ttaaggctag agtacttaat acgactcact 1080
ataggctagc ctcgagaatt cacgcgtcga gcatgcatct agggcgggcca attccgcccc 1140
tctccctccc cccccctaa cgttactggc cgaagccgct tgggaataagg ccggtgtgcg 1200
tttgtctata tgtgattttc caccatattg ccgtcttttg gcaatgtgag ggcccgga 1260
cctggccctg tcttcttgac gagcattcct aggggtcttt cccctctcgc caaaggaatg 1320
caaggtctgt tgaattgtcg gaaggaagca gttcctctgg aagcttcttg aagacaaa 1380
acgtctgtag cgaccctttg caggcagcgg aacccccac ctggcgacag gtgcctctgc 1440
ggccaaaagc cacgtgtata agatacacct gcaaaaggcg cacaacccca gtgccacgtt 1500
gtgatttgga tagttgtgga aagagtcaaa tggctctcct caagcgtatt caacaagggg 1560
ctgaaggatg accagaaggc acccatttgt atgggatctg atctggggcc tcggtgcaca 1620
tgctttacat gtgttttagt gaggttaaaa aaacgtctag gccccccgaa ccacggggac 1680
gtggttttcc ttgaaaaaac acgatgataa gcttgccaca acccaccatg aaaacattta 1740
acattttctc acaagatcta gaattagtag aagtagcgac agagaagatt acaatgcttt 1800
atgagataaa taaacatcat gtgggagcgg caattcgta agaaaacagga gaaatcattt 1860
cggcagtaaa tattgaagcg tatataggac gagtaactgt ttgtgcagaa gccattgcga 1920
ttggtagtgc agtttcgaat ggacaaaagg attttgacac gattgtagct gttagacacc 1980
cttattctga cgaagtagat agaagtattc gagtggtaag tccttggtgg atgtgtaggg 2040
agttgatttc agactatgca ccagattgtt ttgtgttaat agaaatgaat ggcaagttag 2100
tcaaaactac gattgaagaa ctcatccac tcaaatatac ccgaaattaa aagttttacc 2160
ataccaagct tggcggggcg ccgcttccct ttagtgaggg ttaatgcttc gagcagacat 2220
gataagatac attgatgagt ttggacaaac atgcactaga atgcagtga aaaaatgctt 2280
tatttgtgaa atttgtgatg ctattgcttt atttgaacc attataagct gcaataaaca 2340
agttaacaac aacaattgca ttcattttat gtttcaggtt cagggggaga tgtggggagg 2400
tttttaaaagc aagtaaaacc tctacaaatg tggtaaaatc cgataaggat cgatccgggc 2460
tggcgtataa gcgaagaggc ccgcaccgat cgcccttccc aacagttgcg cagcctgaat 2520
ggcgaatgga cgcgcctgtg agcggcgcgt taagcgcggc ggggtgtggtg gttacgcgca 2580
gcgtgacccg ttctgcgggc tttcccgctc agcctctaaa ttcgcttttc cctttagggt 2700
ttctgcggcc agctttacgg cacctcgacc gcaaaaaact tgatttgggt gatggttcc 2760
gtagtggggc atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct 2820
ttaatagtgg actcttggtc caaactggaa caaactcaa ccctatctcg gtctattctt 2880
ttgatttata agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgatttaac 2940
aaatatttaa cgcgaatttt aacaaaatat taacgtttac aatttgcctt gatgcggtat 3000
tttctcctta aacatctgtg cggatattca caccgcatac gcggatctgc gcagcaccat 3060
ggcctgaaat accctctgaa agaggaaact ggttaggtac cttctgaggc ggaaagaacc 3120
agctgtggaa tgtgtgtcag ttaggggtgt gaaagtcccc agggctcccc gcaggcagaa 3180
gtatgcaaag catgcatctc aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3240
cagcaggcag aagtagcaaa agcatgcacg tcaattagtc agcaaccata gtcccggccc 3300
taactccgcc catcccggcc ctaactccgc ccagttccgc ccattctcgc ccccatgggt 3360
gactaatttt ttttatttat gcagaggccg aggcgcctc ggcctctgag ctattccaga 3420
agtagtgagg aggccttttt ggaggcctag gcttttgcaa aaagcttgat tcttctgaca 3480
    
```

-104-

caacagtctc	gaacttaagg	ctagagccac	catgattgaa	caagatggat	tgacgcagg	3540
ttctccggcc	gcttgggtgg	agaggctatt	cggctatgac	tgggcacaa	agacaatcgg	3600
ctgctctgat	gccgcgctgt	tcgggctgtc	agcgcagggg	cgcccgggtc	tttttgtcaa	3660
gaccgacctg	tcgggtgccc	tgaatgaact	gcaggacgag	gcagcgcggc	tatcgtggct	3720
ggccacgacg	ggcggttcctt	gcgcagctgt	gctcgacgtt	gtcactgaag	cggggaaggga	3780
ctggctgcta	ttgggcgaag	tgccggggca	ggatctcctg	tcactctacc	ttgctcctgc	3840
cgagaaagta	tccatcatgg	ctgatgcaat	gcggcggtgt	catacgcttg	atccggctac	3900
ctgcccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	ggatggaagc	3960
cggctctgtc	gatcaggatg	atctggacga	agagcatcag	gggctcgcgc	cagccgaact	4020
gttcgccagg	ctcaaggcgc	gcatgcccga	cggcgaggat	ctcgtcgtga	cccatggcga	4080
tgcttgcttg	ccgaatatca	tggtggaaaa	tggccgcttt	tctggattca	tcgactgtgg	4140
ccggctgggt	gtggcggaac	gctatcagga	ctagcggttg	gctaccgtg	atattgctga	4200
agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggtatcg	ccgctcccga	4260
ttcgcagcgc	atcgcttctc	atcgcttctc	tgacgagttc	ttctgagcgg	gactctgggg	4320
ttcgaaaagta	ccgaccaagc	gacgcccac	gatggccgca	gatggccgca	ataaaatc	4380
tttattttca	ttacatctgt	gtgttggttt	tttgtgtgaa	tcgatagcga	taaggatccg	4440
cgtatgggtg	actctcagta	caatctgctc	tgatgccgca	tagttaagcc	agccccgaca	4500
cccgccaaca	cccgctgacg	cgccctgacg	ggcttgtctg	ctcccggcat	ccgcttacag	4560
acaagctgtg	accgtctccg	ggagctgcat	gtgtcagagg	ttttcacctg	catcacccga	4620
acgcgcgaga	cgaaagggcc	tcgtgatacg	cctattttta	taggttaatg	tcatagataat	4680
aatgggttct	tagactctcag	gtggcacttt	tcgggggaaat	gtgcgcggaa	cccctattttg	4740
tttatttttt	taaatcat	caaatatgta	tcgctcatg	agacaataac	cctgataaat	4800
gcttcaataa	tattgaaaaa	ggaagagtat	gagtatccaa	catttccgtg	tcgcccttat	4860
tccttttttt	gcggcatttt	gccttcctgt	ttttgctcac	ccagaaaacgc	tggtgaaagt	4920
aaaagatgct	gaagatcagt	tggtgacacg	agtgggttac	atcgaaactgg	atctcaacag	4980
cggtaagatc	cttgagagt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	5040
agttctgcta	tgtggcgcg	tattatccc	tattgacgcc	gggcaagagc	aactcggtcg	5100
ccgcatacac	tattctcaga	atgacttggt	tgagtactca	ccagtcacag	aaaagcatct	5160
tacggatggc	atgacagtaa	gagaattatg	cagtgtcgcc	ataaccatga	gtgataaac	5220
tgcgccaac	ttacttctga	caacgatcgg	aggacccaag	gagctaaccg	cttttttgca	5280
caacatgggg	gatcatgtaa	ctcgcttga	tcgttgggaa	ccggagctga	atgaagccat	5340
accaaacgac	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaat	5400
attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	5460
ggataaagtt	gcaggaccac	ttctgcgctc	ggcccttcg	gctggctgg	ttattgctga	5520
taaatctgga	gccggtgagc	gtgggtctcg	cggtatcatt	gcagcactgg	ggccagatgg	5580
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggtgaaacg	5640
aaatagacag	atcgctgaga	taggtgcctc	actgattaa	cattggtaac	tgtagacca	5700
agtttactca	tatatacttt	agattgat	aaaacttcat	ttttaattta	aaaggatcta	5760
ggtgaagatc	ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcggtcca	5820
ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	ttttctcg	5880
cgtaatctgc	tgcttgcaaa	caaaaaaacc	accgctacca	gcgggtggtt	gtttgccgga	5940
tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	6000
tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcacccgc	6060
tacataacct	gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtctg	6120
tcttaccggg	ttggactcaa	gacgatagtt	accggataag	gcgcagcgg	cgggctgaac	6180
gggggggttcg	tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	6240
acagcgtgag	ctatgagaaa	gcgccacgct	tcccgaagg	agaaaggcgg	acaggatatcc	6300
ggtaagcggc	agggtcggaa	caggagagcg	cacgaggagg	cttcagggg	gaaacgcctg	6360
gtatctttat	agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	6420
ctcgtcaggg	ggcgaggacc	tatggaaaaa	cgccagcaac	gcggcctttt	tacgggtcct	6480
ggccttttgc	tgcccttttg	ctcacatggc	tcgacagatc	t		6521

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/85,87,90; A01N 43/34; C12N 15/09

US CL : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS (EAST); STN (MEDLINE, BIOSIS, CAPLUS)**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 01/07572 A2 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 01 February 2001 (01.02.2001), see entire document.	1,2,7,8,11,23-25,27,50 ----- 3-6,9,10,12,13,26,28-
X --- Y	WO 00/11155 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 02 March 2000 (02.03.2000), see entire document.	1,2,7,8,11,23-25,27,50 ----- 3-6,9,10,12,13,26,28-
X --- Y	US 6,171,861 B1 (HARTLEY et al.) 09 January 2001 (09.01.2001), see entire document, especially sequence listing.	1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85 ----- 4-7,9,13-22,28-30,32,34,35,37,38,41,42,46-49,51-53,56,67-71,84,86



Further documents are listed in the continuation of Box C.



See patent family annex.

Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"B" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

01 November 2002 (01.11.2002)

Date of mailing of the international search report

27 NOV 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks

Box PCT

Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Daniel M Sullivan

Telephone No. 703-308-0196

INTERNATIONAL SEARCH REPORT

PCT/US02/17452

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 96/40724 A1 (LIFE TECHNOLOGIES, INC.) 19 December 1996 (19.12.96), see entire document, especially SEQ ID NO: 1-16.	1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85 ----- 1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85
Y	WO 94/00569 A1 (GENPHARM INTERNATIONAL, INC.) 06 January 1994 (06.01.94), see entire document.	72-78
Y	WO 94/23049 A2 (THE JOHNS HOPKINS UNIVERSITY) 13 October 1994 (13.10.94), see entire document.	72-78
X --- Y	US 5,721,118 A (SCHEFFLER) 24 February 1998 (24.02.98), see entire document, especially Example 3.	14,15,22-25,57 ----- 16-21,26,58-64,72-78,84,85,98-105,123
X --- Y	US 5,948,653 A (PATI et al.) 07 September 1999 (07.09.99), see entire document, especially the first full paragraph in column 29.	14,15,22-25,57 ----- 16-21,26,58-64,72-78,84,85,98-105,123
X --- Y	GB 2 331 752 A (MEDICAL RESEARCH COUNCIL) 02 June 1999 (02.06.99), see entire document, especially pages 8-10.	109,123 ----- 72-78,110-122
A	HALDIMANN et al. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol. November 2001, Vol. 183 No. 21, pages 6384-6393, see entire document.	65,66
X,P --- Y,P	MORALLI et al. Insertion of a loxP site in a size-reduced human accessory chromosome. Cytogenet. Cell Genet. 2001, Vol. 94 No. 3-4, pages 113-120, see entire document.	91-93,106,108 ----- 94-97,107
X --- Y	LORBACH et al. Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. J. Mol. Biol. March 2000, Vol. 296 No. 5, pages 1175-1181, see entire document.	1,2,8,50 ----- 3-7,9-13,23-32,51-56,84,85
X --- Y	CALL et al. A cre-lox recombination system for the targeted integration of circular yeast artificial chromosomes into embryonic stem cells. Hum. Mol. Genet. July 2000, Vol. 9, No. 12, pages 1745-1751, see entire document.	14,15,22-26,57,59 ----- 16-21,58,60,84,85,98-105
Y,P	HADLACZKY, G. Satellite DNA-based artificial chromosomes for use in gene therapy. Curr. Opin. Mol. Ther. April 2001, Vol. 3 No. 2, pages 125-132, see entire document.	1-123

INTERNATIONAL SEARCH REPORT

PCT/US02/17452

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-64, 67-71, 79, 84-96, 91-108, and 123, drawn to eukaryotic recombinogenic chromosomes used for introducing heterologous nucleic acids into a chromosome and the resulting cells.
Group II, claim(s) 65-66 and 87-89, drawn to a lambda intR mutein.
Group III, claim(s) 72-78, drawn to the production of transgenic animals.
Group IV, claim(s) 80, drawn to the production of an artificial chromosome library.
Group V, claim(s) 81-83, drawn to a library of cells for genomic screening.
Group VI, claim(s) 89-90, drawn to a modified iron-induced promoter.
Group VII, claim(s) 109-122, drawn to a method for screening compounds and their effects on regulatory regions.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of Group I which defines an advance over the art is a eukaryotic chromosome containing recombinogenic sites that can be used to introduce heterologous nucleic acids into chromosomes, and cells containing these recombinogenic chromosomes.

The special technical feature of Group II involves a lambda intR mutein. This feature defines an advance over Group I in that it involves a protein that is not required for the technical features as set forth above in Group I.

The special technical feature of Group III involves the production of a transgenic animal, which represents a second method for using the invention as set forth above in Group I.

The special technical feature of Group IV involves the production of artificial chromosome expression system libraries, which represents a third method for using the invention as set forth above in Group I.

The special technical feature of Group V involves a library of cells containing the artificial chromosome expression system libraries set forth above in Group IV, and represents a first product resulting from Group IV.

The special technical feature of Group VI involves a modified iron-inducible promoter. This feature defines an advance over Group I in that it involves a promoter that is not required for the technical features as set forth above in Group I.

The special technical feature of Group VII involves a method for screening compounds for their effect on regulatory regions, which represents a fourth method of using the invention as set forth above in Group I.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.